THIS WEEK

The wrong kind of carbon cut

The repeal of Australia's carbon-pricing scheme — the first time a nation has reversed action on climate change — sets a worrying example for other countries mulling steps to reduce emissions.

Climate politics tends to frustrate those convinced that action is needed to halt global warming, but 2014 has seen the stirrings of optimism once again. The European Union (EU) and the United States have both announced concrete plans to curb emissions beyond 2020, and China has signalled a growing willingness to tackle pollution seriously at home and find ways to take part in a binding international climate agreement scheduled to be signed in Paris next year.

So how much does Australia's decision to scrap its carbon-pricing scheme after just two years matter? As we report on page 392, the move was expected and may not directly affect the odds of reaching a meaningful international agreement in 2015 — but it does threaten to undermine the political momentum that has been building up and that must be sustained if the world is to get its act together on the key environmental issue of our time.

A growing body of research indicates that parts of Australia, already plagued by frequent heat, drought and floods, are excessively vulnerable to climate change and its effect on extreme weather. But the nation has also gained symbolic significance in the global debate (and dispute) over the policy and economics of climate change. Climate issues tend to have an enormous influence on domestic policy-making in Australia — from the 1997 Kyoto Protocol (Australia's late signature helped to push it into effect) and the ill-fated 2009 Copenhagen climate summit, to the ongoing debate over cap-and-trade and carbon pricing. Discourse and political battles Down Under can serve as a preview of what is coming up in the greater climate-policy arena.

So are carbon taxes and emissions-trading schemes doomed? Certainly not. Economists have consistently recommended carbon pricing as the most effective way to tackle climate change. And the market-based mechanisms that have been established in Europe, North America and, increasingly, Asia, have by and large helped to control emissions and to stimulate innovation of clean-energy technologies.

Some schemes certainly function better than others. The EU's mandatory emissions-trading system for about 12,000 large industrial emitters is hampered by loopholes and market failures that prevent it from being more efficient. But with reforms under way — emissions allowances are increasingly being auctioned rather than allocated free of charge — the system will remain the cornerstone of EU climate policy.

Australia's scheme, which involved a mere 350 companies, also proved relatively effective in reducing emissions during its short life. Its repeal, and the political circumstances that spurred the move, might seem a bizarre tale of missed opportunity.

Champions of carbon taxes tend to ignore that although people in democratic societies are willing to pay for environmental policies, they are willing to pay only so much. Prime Minister Tony Abbott has politically capitalized on that iron law, rather than taking a risk on telling voters that the modest price they did pay for the scheme is a much-needed investment in the future of their children. But politics is not rational, and future generations are not today's voters. Australia's decision highlights the crude reality that political decision-making does not necessarily follow the logics of science and economics. Politics tends to follow short-term consumer preferences that are all too easily influenced and confused by the power and money of carbon producers. Australia, one of the world's richest countries and a leading per capita carbon emitter, must do more to reassure the rest of the world that it is ready to join

> *"Political decision-making does not necessarily follow the logics of science and economics."*

global efforts to tackle climate change. If it does not, it would send a devastating signal to rising economies such as China that will play a key part in negotiations leading up to an international climate agreement. These nations must be involved in global climate action, but they will hardly be persuaded to sign up if an industrial power the size of Australia falters.

The opportunity for the Australian government to prove its determination is not far off: by April 2015, some nations including Australia are supposed to lay on the negotiating table their offers for the climate conference in Paris in December next year, so that others have enough time to scrutinize the numbers. Australia's current emissions-reduction goal — 5% below 2000 levels by 2020 — is not just modest, it is an affront.

Abbott has won a political battle at home — but to be taken seriously abroad, he must show that he wants to serve his nation more than he wants to appease the fossil-fuel industry. ∎

Fusion furore

Soaring construction costs for ITER are jeopardizing alternative fusion projects.

Fusion energy promises to combine the benefits of renewable resources — clean, carbon-free electric power — with the best qualities of fossil fuels: power day and night, without regard for the vagaries of weather.

The reality is much messier. Fusion power demands heating certain isotopes of hydrogen or other light elements to hundreds of millions of kelvin until they form ionized plasma. The plasma is contained by magnetic fields in a toroidal (doughnut-shaped) chamber until the nuclei fuse and convert mass into energy.

Physicists have struggled to harness fusion for more than six decades. Only in 2006 did an international consortium sign an agreement to start work on ITER, the first reactor designed to 'ignite' fusion plasma such that it will be able to sustain its burn and generate

Chile needs better science governance and support

Its scientists have much to offer the world, but are being held back by scattered administration and changing policies, argues **Pablo Astudillo Besnier**.

With the World Cup over and football fans home from Brazil, the international spotlight will surely fade on South America once more. The continent is changing rapidly, but 1,000 kilometres southwest of Brazil, researchers in Chile would settle for just one step forward: a governance that promotes scientific development and a greater focus on science policy within the government.

Chile may seem distant from other nations when it comes to science. Yet, as countries across the world grapple with the problem of how to restructure their priorities in the wake of the financial crisis, Chile offers a warning about how science and innovation can stagnate without clear governance.

At first glance, science in Chile can seem strong. The country can be rightly proud of its world-class astronomers — highlighted in an article (see *Nature* **510,** 204–206; 2014) in *Nature*'s special on South America last month — and, as the government likes to point out, we are a highly productive player in Latin American science, ranking above Brazil, Mexico and Argentina in terms of the number of articles per capita.

Yet, despite the glossy images of the telescopes promising a high-tech future, in 2012, Chile spent just 0.35% of its gross domestic product on research and development (R&D) — the least of all countries in the Organisation for Economic Co-operation and Development. Two-thirds of the academic publications come from just five universities. And, perhaps most importantly, a heavy focus on applied science threatens to stifle basic research and its potential to innovate.

Put simply, science does not have a strong voice in Chile. Administration of science and research is scattered across government ministries. Without a ministry to coordinate and implement a national agenda, responsibilities have been unclear and policies are subject to disruptions and cancellations. As a result, opportunities for scientists to develop momentum are lost, as is a lot of quality work. As responsibility for science drifted through the hands of a variety of government officials, most tried to push the scientific agenda towards the needs of the industrial sector, disregarding the academic, political and social roles that science must also have.

This chaotic governance mirrors the broader political turmoil that the country has endured in recent decades. Our main funding body — the National Commission for Scientific and Technological Research — was created in 1967, perhaps one of the few moments in which the Chilean government saw science as a clear priority. However, political changes in the subsequent years ended its role as an advisory council, and it never recovered.

In 2005, the authorities did acknowledge that the governance of science and innovation was deficient, and tried to give it a more defined structure, creating the National Innovation Council for Competitiveness and then a new interministerial committee. Following the advice of this council, and with the aim of using science to boost economic growth and competitiveness, the committee prioritized funds to promote clusters of research — including on processed foods, the mining and farming industries, and financial services. In this setting, scientists were never clearly heard.

Support for basic R&D has been inconsistent. The number of grants awarded by the National Fund for the Scientific and Technological Development (the main basic-research funding source) fell from a high of more than 500 in 1991 to around 350 between 2000 and 2005, but rose again after that.

The government elected in 2010 abandoned the cluster policy; its successor, which came into power earlier this year, has pledged to restart it. As well as making it hard to do quality research, the unsystematic approach means that projects and policies rarely run for long enough to be evaluated. As such, we do not know what impact — if any — they have had.

What we do know is that Chile's ranking in terms of innovation has fallen persistently in the past seven years. International reports confirm that the nation is overly reliant on the export of commodities.

Programmes related to R&D have been created without careful planning. An example is the scholarship programme for graduate studies both in Chile and abroad. Between 2008 and 2012, the government awarded more than 8,000 postgraduate scholarships. But fewer than 200 awards were granted to incorporate scientists into the academic and industrial sectors. Where will all the newly qualified scientists work?

This is the conundrum of science in Chile.

Last year, we thought we had achieved a step forward. Outgoing President Sebastián Piñera had finally pledged to create a science and innovation ministry. Yet the new government has put the scheme on hold, and talks instead about strengthening the role of the ministry of the economy — an idea that lacks support and empirical evidence, and one that does little to increase the contribution of science to relevant public issues. Research and innovation are of course important for the economy and competitiveness, but that should not be the only focus of a science policy.

In Brazil, Chile's footballers impressed the watching world. Its scientists can do the same, but they need someone at the top to give them the chance. ■

> **CHILE OFFERS A WARNING ABOUT HOW SCIENCE AND INNOVATION CAN STAGNATE WITHOUT CLEAR GOVERNANCE.**

Pablo Astudillo Besnier *is a postdoctoral fellow with the Becas Chile programme, and founder of the More Science Foundation in Santiago. e-mail: pastudillo@fmasciencia.cl*

MICROBIOLOGY

A cheesy tool for the laboratory

Cheese rinds could help to reveal how microbial communities form and species interact.

Microbial communities affect ecosystems and human health, but are difficult to study in the lab. To find microbial systems that can be easily manipulated, Rachel Dutton and her colleagues at Harvard University in Cambridge, Massachusetts, studied rind samples from more than 100 types of cheese, including Brie and Camembert.

They found that many of the bacterial and fungal species that grow on ageing cheese are easily cultured. The team used sequencing to identify key interactions between bacteria and fungi and to track the development of the microbial community on a cheese as it aged. Moreover, the researchers could reconstruct many of these interactions *in vitro*.

Studying cheese rinds could provide insight into other microbial communities, such as those found on skin, the authors say.
Cell 158, **422–433 (2014)**

CLIMATE CHANGE

Hotter summers despite hiatus

Rising greenhouse-gas levels have been making summers in the Northern Hemisphere hotter, even though global warming has been slowing in recent years.

Youichi Kamae of the National Institute for Environmental Studies in Tsukuba, Japan, and his colleagues compared the results of climate models that include the effects of human

activity with those that do not.

They found that a rise in greenhouse-gas concentrations in the atmosphere has been the dominant cause of the increasing frequency of unusually hot summers in the

PLANETARY SCIENCE

Frost forms Mars gullies

Gullies on Mars were probably not created by liquid water but by the seasonal freezing and thawing of carbon dioxide, according to an analysis of high-resolution images.

Many scientists have argued that flowing water — a prerequisite for life — carved the gullies (pictured) that are widespread across Mars. Colin Dundas of the US Geological Survey in Flagstaff, Arizona, and his team studied 98 gully sites in Mars's northern hemisphere and 258 in the southern half, using data from NASA's Mars Reconnaissance Orbiter.

Looking at the same areas each year, the researchers saw the gullies growing and changing shape at the same time as carbon dioxide frost appeared and disappeared. Liquid water is not required in this process, the authors say.
Icarus http://doi.org/trx (2014)

Northern Hemisphere since the late twentieth century. In the middle latitudes, however, about half of the increase in hot summers can be attributed to natural climate variability over the Pacific and Atlantic oceans.

Heat extremes over land will probably become more frequent even if the global-warming hiatus persists, the authors conclude.
Geophys. Res. Lett. http://doi.org/tr3 (2014)

PALAEONTOLOGY

Foraging patterns found in fossils

Researchers have discovered the first fossil evidence of a type of search behaviour displayed by some modern animals when looking for food.

Animals that hunt sparse prey over large areas often move in patterns known as Lévy walks — characterized by numerous small steps interspersed with rare long jumps to optimize foraging. A team led by David Sims at the Marine Biological Association in Plymouth, UK, discovered the pattern in 50-million-year-old fossilized tracks made by an extinct sea urchin, *Scolicia*. Its movements resembled those of albatrosses searching for squid over an open ocean.

The sea urchins may have evolved this foraging strategy after global resource collapses made their food supplies sparse, the researchers say.
Proc. Natl Acad. Sci. USA http://doi.org/trs (2014)

BIOTECHNOLOGY

Gene edits boost wheat defences

Researchers have used advanced gene-editing techniques to generate disease-resistant wheat.

Genetically altering *Triticum aestivum* wheat is difficult to do, in part because the plant has six sets of chromosomes instead of the two sets found in humans. So Caixia Gao and

Jin-Long Qiu of the Chinese Academy of Sciences in Beijing and their colleagues used two gene-editing approaches — TALEN enzymes and the CRISPR–Cas9 system — to disable a gene called *MLO* in all of the plants' chromosomes. This made the plants resistant to fungal diseases called powdery mildew. Knocking out all versions of the gene yielded the greatest resistance.

For the many crops that have multiple genome copies, such techniques can lead to improvements that are not possible through conventional breeding, the authors suggest.

Nature Biotechnol. http://dx.doi.org/10.1038/nbt.2969 (2014)

ZOOLOGY

Mystery bird is dodo relative

A dead pigeon specimen that has lain for years in a UK museum has been confirmed by DNA analysis as a new species — and as a relative of the dodo.

Tim Heupink of Griffith University in Brisbane, Australia, and his colleagues extracted and sequenced very short DNA fragments from the only remaining specimen of the spotted green pigeon (*Caloenas maculata*; artist's impression **pictured**). After being described in 1783, it ended up in a museum in Liverpool, UK, but nothing else was known about it.

Some researchers had claimed that the specimen was merely a Nicobar pigeon (*Caloenas nicobarica*), but the authors determined that *C. maculata* is a separate

species and that both birds share an ancestor with the dodo (*Raphus cucullatus*). This ancestor was probably a semi-terrestrial bird that island-hopped from southeast Asia or India across the oceans, eventually evolving into the dodo and other pigeon species that live on remote islands.

BMC Evol. Biol. 14, **136** (2014)

MICROBIOLOGY

Injuries invite ulcer microbe

A microbe that can lead to the formation of stomach ulcers and cancer quickly finds its way to tiny injuries in the stomach lining and colonizes them, slowing healing.

Marshall Montrose at the University of Cincinnati in Ohio and his colleagues exposed mice with stomach injuries to *Helicobacter pylori* and found that the damaged sites had larger colonies of the microbe than healthy areas.

Bacterial strains that had been engineered to be immobile or to be insensitive to their environment were less able to infect wounds than were normal strains. Moreover, the authors found that the bacterium takes only a few minutes to navigate from elsewhere in the stomach to damaged areas to slow repair.

They suggest that even microscopic injuries in the stomach that occur through eating and other normal activities are vulnerable to infection.

PLoS Pathogens 10, **e1004275** (2014)

ASTROPHYSICS

Radio burst from beyond the Galaxy

A telescope has detected a mysterious millisecond burst of radio waves that seems to be coming from outside the Milky Way.

Laura Spitler at the Max Planck Institute for Radio Astronomy in Bonn, Germany, and her colleagues found the burst using the Arecibo

radio telescope in Puerto Rico. The Parkes telescope in New South Wales, Australia, had previously picked up similar pulses — matching in brightness and duration — but a lack of comparable findings from other instruments at the time led astronomers to speculate that the signals were caused by instrument error or by radio interference from human sources.

Possible origins of the pulses include evaporating black holes, mergers of neutron stars or flares from magnetically active stars, say the authors.

Astrophys. J. 790, **101–109** (2014)

ARCHAEOLOGY

Clovis people were hunters in Mexico

Elephant-like animals called gomphotheres (*Cuvieronius* sp.), thought to have gone extinct long before humans arrived in the Americas, might have stuck around long enough to be hunted by prehistoric people.

At a site called El Fin del Mundo in Sonora, Mexico, a team led by Guadalupe Sanchez of the National Autonomous

Spotlight falls on top 1% in science

An analysis led by John Ioannidis, a health-policy researcher at Stanford University, found that less than 1% of all researchers managed to publish every year from 1996 to 2011, but that those elite few were authors on more than 41% of all papers in the same period. Many noted the similarity between this and claims that the top 1% of US earners hold an inordinate share of the country's wealth. "Occupy!" tweeted Karen James, a geneticist at MDI Biological Laboratory in Maine, alluding to the Occupy Wall Street movement that calls for economic equality. Chris Cramer, a chemist at the University of Minnesota in Minneapolis, tweeted that it was "an interesting example of the top 1% CONTRIBUTING 41% (instead of owning?)."

PLoS ONE 9, **e101698** (2014)

Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

↻ NATURE.COM
For more on popular papers:
go.nature.com/f1c2bn

University of Mexico in Sonora discovered gomphothere bones (jawbone **pictured**) intermingled with stone spear points from 13,400 years ago. These stone tools were made by some of the earliest people to inhabit North America, a group known as the Clovis people. Archaeologists knew that Clovis hunters pursued mammoths and mastodons, but this discovery adds gomphotheres to their diet.

The finding simultaneously extends the period during which these animals were alive and makes the Mexican site one of the oldest and southernmost Clovis sites known.

Proc. Natl Acad. Sci. USA http://doi.org/tr4 (2014)

↻ NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

Microsoft search

Microsoft announced last week that it is to revamp its free scholarly search engine, Microsoft Academic Search. The company plans to stop running the service as a stand-alone product, instead ensuring that information about scholarly works and academics appears through its general search engine, Bing. Microsoft Academic Search once competed with Google Scholar, but failed to stay up to date (see go.nature. com/ti3t9x). Microsoft says that researchers will be able to control what information about themselves others can access through Bing.

Water savings

Developing nations could save a significant amount of water by adopting modern farming techniques and technologies, says a study published on 15 July by the United Nations Environment Programme. The study estimates that the Indian state of Punjab could reduce annual agricultural water use by up to 23 billion cubic metres by changing crop patterns and using water-saving technologies. The state could replace rice with a mixture of crops, including pulses, maize (corn), cotton, fruit, vegetables and sugar cane. It could also sow rice rather than plant seedlings.

Burnout journal

The publisher Elsevier launched an open-access research journal focused on stress-related workplace fatigue on 15 July. An editorial in the first issue of *Burnout Research* states that burnout is a "longstanding issue of social concern" that affects both job performance and quality of life. The editors say that 1,000 articles are published on

High risk

Low risk
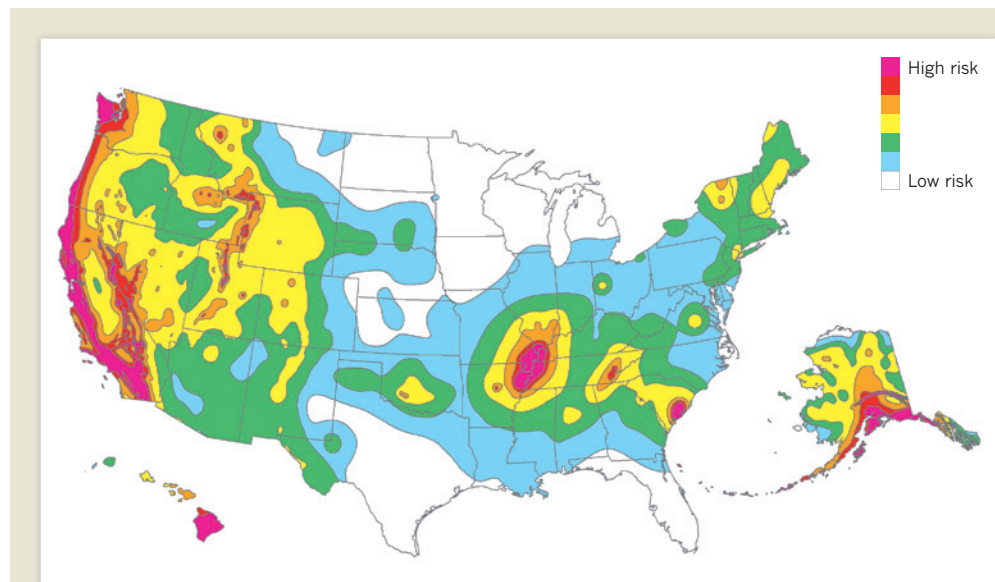
USGS

Fresh look at US earthquake risk

Seismic-hazard maps for the United States released on 17 July suggest that some regions are at greater risk of earthquake damage than they were in 2008. The US Geological Survey has produced an update that takes into account the latest methods, models and data.

A magnitude-5.8 earthquake in Virginia in 2011 has shown that the eastern United States has increased potential for large, damaging earthquakes. And New York City is slightly less likely than before to be hit by slow shaking, the type of shaking that damages tall buildings.

burnout every year, scattered across 100 journals; they hope that their journal will help to bring coherence to the field.

Quake alert system

A US congressional appropriations committee approved US$5 million on 15 July to develop an earthquake early-warning system for the US West Coast. It will monitor geological threats such as the San Andreas and Cascadia fault zones in California, Oregon and Washington. The system would alert locations such as hospitals and transportation centres that an earthquake has begun, providing crucial seconds of warning before ground-shaking starts. The bill must still be approved by both

houses of Congress. Other countries, including Japan and Mexico, already have such warning systems.

Broad donation

The Broad Institute in Cambridge, Massachusetts, has been given what it describes as the largest-ever donation for psychiatric research. On 22 July, the institute announced that philanthropist Ted Stanley will donate US$650 million to its Stanley Center for Psychiatric Research, established in 2007. See page 393 for more.

HIV experts killed

Leading HIV researcher Joep Lange and five other delegates to a major AIDS meeting are among the 298 people who

died on board the Malaysia Airlines flight MH17, which crashed on 17 July. They were on their way to the 20th International AIDS Conference being held on 20–25 July in Melbourne, Australia. The plane was flying from Amsterdam to Kuala Lumpur when it was shot down by a surface-to-air missile over eastern Ukraine, killing all on board. See page 391 for more.

Chikungunya cases

Florida's health department announced the first locally acquired cases of the mosquito-borne chikungunya virus in the United States on 17 July. The two cases confirm that the virus has infected US mosquitoes; previous US cases were acquired abroad. First identified in East Africa

in 1952, it appeared in the Western Hemisphere on the island of St Martin in December 2013 and quickly spread to 22 other nations in the Americas. It causes rashes and joint pain that can persist for years. See go.nature.com/5bv5el for more.

Pathogens found

The US Food and Drug Administration revealed on 16 July that forgotten vials of smallpox found this month at the campus of the National Institutes of Health in Bethesda, Maryland, were not alone. A total of 12 boxes and 327 vials, believed to be more than 50 years old, also included vials labelled as dengue, influenza, Q fever and rickettsia. There is no evidence that anyone was exposed to any of the pathogens, the agency adds.

Binary comet

The comet-chasing spacecraft Rosetta has taken images revealing its target's surprising shape. Pictures published on 17 July show that 67P/Churyumov-Gerasimenko consists of two distinct parts (**pictured**), known as a contact binary, and may have formed when separate comets collided at low speed or a single comet was pulled apart or eroded. The European Space Agency probe is scheduled to rendezvous with its target on

6 August, then study it in detail before setting down a lander in November. That may be made more difficult by the comet's irregular shape. See go.nature.com/a1xghw for more.

Darwin library

Science historian John van Wyhe of the National University of Singapore has released a virtual version of Charles Darwin's library from the naturalist's famous 1830s journey around the world on HMS *Beagle*. For the first time, all of the printed sources that Darwin referenced during the voyage are available online. The library, part of a broader project called Darwin Online, comprises 404 volumes, totalling more than 195,000 pages. See go.nature.com/znwlnn for more.

Carbon tax dropped

The Australian Senate last week voted 39 to 32 to scrap the country's pioneering

carbon-pricing scheme, established just two years ago with the goal of slashing greenhouse-gas emissions. The scheme, second only to the European Union's cap-and-trade scheme in terms of the quantity of emissions it aimed to control, will be replaced with an incentive-based system to encourage farmers and companies to reduce emissions and restore soil carbon. Experts fear that this approach will be less efficient than a market-based mechanism. See pages 383 and 392 for more.

CDC reforms

The US Centers for Disease Control and Prevention is taking steps to create a "culture of lab safety" in response to incidents this year in which H5N1 influenza and anthrax were mishandled, director Thomas Frieden told the US Congress's oversight and investigations subcommittee on 16 July. Frieden said that he plans a broad assessment of the agency's inventory of dangerous agents, to review its training procedures, and to bring in an outside panel to advise on safety. See go.nature.com/bdhkxe for more.

Stem-cell patents

The European Court of Justice seems set to allow patents on some technologies that create stem cells from human eggs. Almost three years ago,

it declared that any patent that relied even indirectly on human embryonic stem-cell lines should not be granted on grounds of immorality (see *Nature* **480,** 310–312; 2011). But on 17 July, Pedro Cruz Villalón, the court's advocate general (whose opinions are usually accepted), said that two patent applications based on creating pluripotent stem cells from non-fertilized human eggs should be allowed because the cells are intrinsically unable to develop into humans. The court's final ruling is expected in a few months.

Obokata PhD probe

Haruko Obokata, the author of two stem-cell papers that were retracted from *Nature* after she was found guilty of scientific misconduct, should not lose her PhD, an investigative panel ruled last week. The independent investigation initiated by Waseda University in Tokyo, which awarded Obokata's PhD in 2011, found six instances of misconduct in her thesis. On 17 July, the panel said in a report that the problems were either unintentional or were not significant enough to influence the acceptance of her thesis. The university is now considering how to take the recommendations forward.

Private spaceflight

The United Kingdom has identified eight coastal airfields as candidate sites for a commercial spaceport planned to open in 2018, the government announced on 15 July. Six of the sites are in Scotland; England and Wales each have one. The previous week, the US Federal Aviation Authority granted approval for the private launch firm SpaceX of Hawthorne, California, to construct a spaceport in southern Texas.
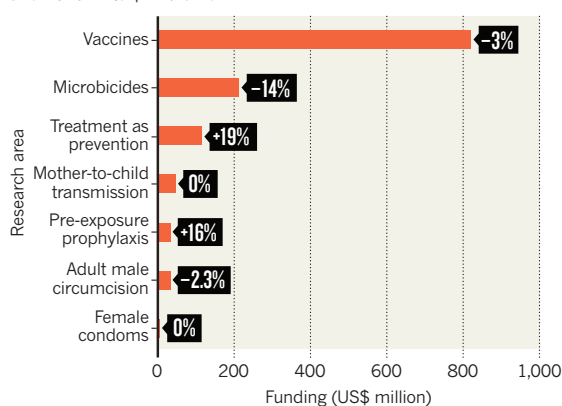
↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

Funding for HIV-prevention research has fallen for the first time in six years, according to the HIV Vaccines and Microbicides Resource Tracking Working Group. Vaccine research remains the major focus. Support has grown for research into pre-exposure prophylaxis (treating people at risk) and treatment of infected people to prevent transmission, but fallen for that into topical microbicides. Work on adult male circumcision also received less funding (partly because it is a proven strategy).

FUNDING FALLS FOR HIV-PREVENTION RESEARCH

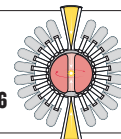Global investment slipped by US$50 million, or 4%, between 2012 and 2013 — to $1.26 billion.

Research area	Funding (US$ million)	Change
Vaccines		−3%
Microbicides		−14%
Treatment as prevention		+19%
Mother-to-child transmission		0%
Pre-exposure prophylaxis		+16%
Adult male circumcision		−2.3%
Female condoms		0%

0   200   400   600   800   1,000
Funding (US$ million)

NEWS IN FOCUS

MARK DADSWELL/REUTERS

Participants in the International AIDS Conference paid tribute to researchers killed in a plane crash.

COMMUNITY

Delegates mourn MH17 passengers

AIDS conference tries to draw inspiration from lost scientists.

BY KATIA MOSKVITCH IN LONDON AND EDWARD SUSMAN IN MELBOURNE, AUSTRALIA

With heads bowed, many hand-in-hand with their neighbours, hundreds of attendees at the 20th International AIDS Conference in Melbourne, Australia, held a moment of silence on 20 July to honour the six HIV researchers and activists who died on Malaysia Airlines flight MH17.

Tributes to the dead scientists, all of whom were on their way to the meeting, were led by 11 former, present and future presidents of the International AIDS Society (IAS), which runs the conference. Françoise Barré-Sinoussi, the current IAS head, dedicated the event to

the delegates — five Dutch and one British — who were killed, and Lambert Grijns, Dutch Ambassador for Sexual and Reproductive Health and Rights and HIV/AIDS, read a letter of condolence. Later, at an exhibition booth run by several Dutch charities, attendees were invited to write notes in a book of remembrance.

"The mood has been one of sadness and loss," Anthony Fauci, director of the US National Institute of Allergy and Infectious Diseases, told *Nature*. "We are feeling the vulnerability for all of us, because there are 12,000 people here, and [most] of us just 2 days ago stepped on a plane to travel thousands of miles to get here, and the tragedy is that some of us didn't make it."

The IAS had earlier confirmed the names of those who died when the plane was shot down over eastern Ukraine on 17 July. They were: Joep Lange, a clinical virologist at the University of Amsterdam and former president of the IAS (see *Nature* http://doi.org/tsx; 2014); his partner Jacqueline van Tongeren, head of communications at the Amsterdam Institute for Global Health and Development, and a long-time AIDS nurse; Glenn Thomas, media-relations coordinator at the World Health Organization in Geneva, Switzerland; Pim de Kuijer, an activist and lobbyist for STOP AIDS NOW! in Amsterdam; Martine de Schutter, a programme manager at Aids Fonds in Amsterdam; and Lucie van Mens, director of programme development at the Female Health Company in Chicago, Illinois, which produces female condoms.

Despite the sombre atmosphere, conference participants were keen to take inspiration from the work and passion of those killed. "Most people are saying that if Lange, Thomas and the others were here, they would want us to continue the struggle," said Fauci. "So it's really a parallel emotion, of sadness and loss but also of commitment to move ahead according to the same passion that the people who are not here have shown over the years."

Sharon Lewin, head of the department of infectious diseases at Monash University in Melbourne, and a co-chair of the conference, said that although sadness lingered, the general mood had changed after the opening ceremony. "There will be further memorials, but in general the conference has returned to its purpose of fighting HIV/AIDS," she told *Nature*.

This year's conference focuses on bringing treatments for HIV and AIDS to people in developing countries. Barré-Sinoussi reminded delegates at the opening session of the need to "step up the pace and redouble our efforts. Too many countries are still struggling to address their HIV epidemic with their most vulnerable people consistently being left behind."

It is in this effort that the loss of Lange, who helped to establish the safety and effectiveness of treatment using multiple antiviral drugs, will be most sorely felt, said Fauci. "He was a very inspiring, passionate catalyst. He was an advocate for getting treatment to whoever needs it, no matter where they are in the world." ∎

Emissions from the Australian energy sector fell by 5% after the introduction of a carbon tax.

POLICY

Anger as Australia dumps carbon tax

Climate experts decry demise of emissions–control system.

BY QUIRIN SCHIERMEIER

Australia's pioneering carbon-pricing mechanism has failed to survive its infancy. In a major victory for Prime Minister Tony Abbott, parliament agreed on 17 July to axe the scheme with immediate effect.

The repeal scraps both the unpopular carbon tax, established in July 2012, and proposals to turn it into a more flexible emissions-trading scheme in mid-2015. The initiative would have seen large companies and utility firms buying and selling emissions allowances in a joint market with the European Union (EU).

The move has disappointed many researchers. "This is a big setback for Australia's climate policy," says Frank Jotzo, a climate economist at the Australian National University in Canberra. "An economically sensible policy framework is being discarded and there is nothing adequate to replace it." He adds that it is unclear how Australia will meet its 2020 emissions-reduction target, "much less how the economy might be put on a trajectory to deeper cuts later on".

In its first year, the carbon tax raised Aus$6.6 billion (US$6.2 billion) from the almost 350 high-polluting companies obliged to pay about Aus$24 for every tonne of carbon dioxide equivalent that they emitted. The tax, introduced by the previous government, led to a 5% decrease in emissions from the power sector.

But critics say that the scheme has vastly increased consumer electricity prices. The government says that abolishing it — a key commitment in Abbott's 2013 election campaign — will reduce the cost of living by an average of Aus$550 per household per year. After the scheme's demise, Abbott described it as a "useless, destructive tax".

Australia's total greenhouse-gas emissions — 538.4 million tonnes of CO_2 equivalent in 2013, down 0.8% from 2012 — account for less than 1.5% of total global emissions. But owing to extensive use of coal in electricity generation, the country is one of the world's largest polluters on a per capita basis.

The carbon-pricing mechanism was expected to reduce the nation's greenhouse-gas emissions by 5% below 2000 levels by 2020. It will be replaced by an Emissions Reduction Fund that will aim to meet the same target by offering financial incentives for companies to increase their energy efficiency, and for land-owners to replenish soil carbon and plant trees. But experts fear that the policy will be more costly and less effective than a market-based emissions-trading scheme.

"The carbon-pricing mechanism has been effective and the proposed alternative of a subsidy-based Emissions Reduction Fund has

many holes," says Jotzo. "For Australia — as one of the richest, highest-emitting and most vulnerable countries to climate change — to repeal the backbone of its mitigation policy sends a negative signal internationally."

The status of other carbon-trading systems is mixed. The EU, which plans to cut its emissions by 40% below 1990 levels by 2030, is currently overhauling its trading scheme to reduce the number of emissions allowances available; the price of carbon has temporarily collapsed as a result of over-allocation of allowances during the recession. China is pushing forward with regional carbon-trading schemes, and the United States — which is reluctant to put a price on emissions — has announced plans to cut emissions from power plants by 30% from 2005 levels by 2030. This will be achieved by improving efficiencies, shifting from coal power to gas and renewables, and implementing regional cap-and-trade emissions schemes.

The United Nations hopes to hammer out a binding international climate agreement by 2015, based on national commitments. As part of this process, major economies including Australia are supposed to submit their post-2020 emissions-reduction targets to the UN by April next year.

But Australia's U-turns on the carbon tax and other climate policies may discourage other nations from being as ambitious as they might have been, says Elliot Diringer, executive vice-president of the US Center for Climate and Energy Solutions in Arlington, Virginia. "Australia's decision underscores the vagaries of domestic policy-making," he says. "It saps rather than strengthens the global momentum — but it will not deter a global climate agreement in 2015."

The Australian Climate Change Authority (CCA), established by the previous government to provide advice on emissions targets, recommended in February that Australia's reduction goal should be 19% below 2000 levels by 2020. The present government intends to close the CCA.

> *"An economically sensible policy framework is being discarded and there is nothing adequate to replace it."*

"The government's relatively weak and ineffective action on climate change is not consistent with a fair contribution from Australia towards global efforts to avoid dangerous climate change," says David Karoly, a climate scientist at the University of Melbourne and a member of the CCA.

Economists think that carbon pricing is the most efficient way to cut emissions. But Michael Grubb, who studies energy and climate policies at University College London, says that this often ignores the political reality — where decision-making tends to follow voters' immediate wishes and concerns. ■ **SEE EDITORIAL P.383**

Gene–hunt gain for mental health

Flood of genetic locations linked to schizophrenia help spark financial boost to research field.

BY SARA REARDON

Researchers seeking to unpick the complex genetic basis of mental disorders such as schizophrenia have taken a huge step towards their goal. A paper[1] published in *Nature* this week ties 108 genetic locations to schizophrenia — most for the first time. The encouraging results come on the same day as a US$650-million donation to expand research into psychiatric conditions.

Philanthropist Ted Stanley gave the money to the Stanley Center for Psychiatric Research at the Broad Institute in Cambridge, Massachusetts. The institute describes the gift as the largest-ever donation for psychiatric research.

"The assurance of a very long life of the centre allows us to take on ambitious long-term projects and intellectual risks," says its director, Steven Hyman.

The centre will use the money to fund genetic studies as well as investigations into the biological pathways involved in conditions such as schizophrenia, autism and bipolar disorder. The research effort will also seek better animal and cell models for mental disorders, and will investigate chemicals that might be developed into drugs.

The *Nature* paper[1] was produced by the Psychiatric Genomics Consortium (PGC) — a collaboration of more than 80 institutions, including the Broad Institute. Hundreds of researchers from the PGC pooled samples from more than 150,000 people, of whom 36,989 had been diagnosed with schizophrenia. This enormous sample size enabled them to spot 108 genetic locations, or loci, where the DNA sequence in people with schizophrenia tends to differ from the sequence in people without the disease. "This paper is in some ways proof that genomics can succeed," Hyman says.

"This is a pretty exciting moment in the history of this field," agrees Thomas Insel, director of the National Institute of Mental Health (NIMH) in Bethesda, Maryland, who was not involved in the study.

Many of the variations seem to be common, so most people will have some of them — but people with schizophrenia have more, and each contributes a small amount to the overall risk of developing the condition. This has made the variants difficult to spot in smaller samples, which is why the PGC collaboration is so crucial, Insel says: "If you want to look for common variants, you have to work with a lot of friends."

Broad population studies are shedding light on the genetic causes of mental disorders.

The large sample size also allowed the researchers to develop an algorithm that would calculate a 'risk score' for each variant's contribution to schizophrenia. This could eventually be used to predict who might develop the disorder or to add weight to an uncertain diagnosis of schizophrenia, Insel says. The NIMH is likely to dedicate more money soon towards doing intensive genetic surveys and following the genetic leads that come out of them, he adds.

Of the 108 loci identified in the PGC paper, 83 had not been identified by previous work. Many of the loci are in or near genes suspected to be involved in schizophrenia and other mental disorders, says Michael O'Donovan, a psychiatrist at Cardiff University, UK, who led the study. One was the gene encoding the dopamine receptor DRD2, which is a target of all current drugs for treating schizophrenia but has never come up as a risk factor. The loci also included genes that encode several proteins involved in transmitting electrical signals between neurons and creating connections between brain cells.

Several loci associated with schizophrenia encode proteins involved in the immune system, which has long been suspected to have a role in triggering the disorder. O'Donovan says that although the link is intriguing, its importance is not yet clear.

"It's a stunning confirmation that there are many, many genetic factors" involved in schizophrenia, says Pamela Sklar, a psychiatric geneticist at Mount Sinai School of Medicine in New York and a co-author of the paper. "There were hints of that from previous work, but this takes it to the next level." O'Donovan says that the PGC plans to double the sample size of people with schizophrenia by next year, which should enable the researchers to pinpoint the most important variants.

The group also plans to expand the data to include DNA from people around the world. The current collection, of nearly 200,000 samples, represents mainly people of northern European descent. "We want to make sure there is global health equity and that we don't develop treatments for just a subset of the world's population," says Hyman.

Schizophrenia is only one of the disorders that the group is working on. A 20 July paper in *Nature Genetics*[2], with many of the same authors as the schizophrenia work, surveyed the genomes of 466 people with autism and some 2,500 people without, and estimated that about 52% of the risk for autism is genetic. Only 2.6% of the risk was linked to mutations that occur spontaneously during development, rather than being inherited. Like schizophrenia, autism seems to result largely from a certain combination of common variants, but people with autism are more likely also to have inherited a smaller number of rare variants that each greatly increase risk, the authors suggest. And because these variants are so rare, a larger sample size will be necessary to determine the truly important ones, says Joseph Buxbaum, a molecular biologist at Mount Sinai, who led the study.

The PGC is also using a platform called PsychChip, which targets parts of the genome associated with several psychiatric illnesses, including some less-studied ones such as anorexia nervosa and obsessive–compulsive disorder. By the end of this year, Sklar says, the team hopes to have sequenced samples from 100,000 people with mental illness: enough to identify the most meaningful associations. ■ **SEE NEWS & VIEWS P.412**

1. Ripke, S. *et al. Nature* http://dx.doi.org/10.1038/nature13595 (2014).
2. Gaugler, T. *et al. Nature Genet.* http://dx.doi.org/10.1038/ng.3039 (2014).

The 27-kilometre Large Hadron Collider at CERN could soon be overtaken as the world's largest particle smasher by a proposed Chinese machine.

PARTICLE PHYSICS

China plans super collider

Proposals for two accelerators could see country become collider capital of the world.

BY ELIZABETH GIBNEY

For decades, Europe and the United States have led the way when it comes to high-energy particle colliders. But a proposal by China that is quietly gathering momentum has raised the possibility that the country could soon position itself at the forefront of particle physics.

Scientists at the Institute of High Energy Physics (IHEP) in Beijing, working with international collaborators, are planning to build a 'Higgs factory' by 2028 — a 52-kilometre underground ring that would smash together electrons and positrons. Collisions of these fundamental particles would allow the Higgs boson to be studied with greater precision than at the much smaller Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

Physicists say that the proposed US\$3-billion machine is within technological grasp and is considered conservative in scope and cost. But

China hopes that it would also be a stepping stone to a next-generation collider — a super proton–proton collider — in the same tunnel.

European and US teams have both shown interest in building their own super collider (see *Nature* **503,** 177; 2013), but the huge amount of research needed before such a machine could be built means that the earliest date either can aim for is 2035. China would like to build its electron–positron collider in the meantime, unaided by international funding if needs be, and follow it up as fast as technologically possible with the super proton collider. Because only one super collider is likely to be built, China's momentum puts it firmly in the driving seat.

Speaking this month at the International Conference on High Energy Physics in Valencia, Spain, IHEP director Yifang Wang said that, to secure government support, China wanted to work towards a more immediate goal than a super collider by 2035. "You can't just talk about a project which is 20 years from now," he said.

Electron–positron colliders and hadron colliders such as the LHC complement each other. Hadron colliders are sledgehammers, smashing together protons (a kind of hadron that comprises three fundamental particles called quarks) at high energies to see what emerges. Lower-energy electron–positron machines produce cleaner collisions that are easier to analyse, because they are already smashing together fundamental particles. By examining in detail the interactions of the Higgs boson with other particles, the proposed Chinese collider should, for example, be able to detect whether the Higgs is a simple particle or something more exotic. This would help physicists to work out whether the particle fits with predictions made by the standard model of particle physics, or whether, for example, multiple types of Higgs boson exist.

The machine would be a big leap for China. The country's biggest current collider is just 240 metres in circumference. Ten years ago, Chinese particle physicists would have doubted

PLANETARY SCIENCE

Mars slow to yield its secrets

Scientists seek fresh approaches to deciphering red planet's history.

**BY ALEXANDRA WITZE
PASADENA, CALIFORNIA**

Thunderous applause greeted planetary cartographer Ken Tanaka of the US Geological Survey on 14 July as he unveiled a new geological map of Mars. A culmination of more than a decade of probing the red planet's geology and history with orbiters and rovers, it revealed that large areas of the Martian surface are more ancient than once thought.

But by the time the Eighth International Conference on Mars closed five days later, the mood was less celebratory and more introspective. Talks and presentations at the meeting in Pasadena, California, had covered major recent discoveries on Mars, and yet scientists found themselves asking the same questions they had been chasing for decades: when did Mars have water on its surface, and what happened to it? Did life ever exist there?

"There's a lot of knowledge and not so much understanding," says Phil Christensen, a Mars geologist at Arizona State University in Tempe. "The devil continues to be in the details."

The sustained programme of Mars exploration pursued by NASA and the European Space Agency has unquestionably yielded a bounty of discoveries. Satellites such as the Mars Reconnaissance Orbiter have watched dust storms envelop the planet and carbon dioxide at its poles freeze and thaw. Rovers such as Curiosity have closely analysed rocks to determine where and when Mars might have contained watery, life-friendly environments.

But a thorough, planet-wide understanding of Martian history remains elusive. "It reminds me of Earth science in the 1950s — lots of observations but no overarching breakthrough yet," says David Des Marais, an astrobiologist at NASA's Ames Research Center in Moffett Field, California.

Among other things, Mars researchers have found it increasingly hard to explain how the planet might have stayed warm and wet in its early history (see *Nature* **484,** 153; 2012). Sprawling river channels and water-rich clay minerals suggest that water once coursed
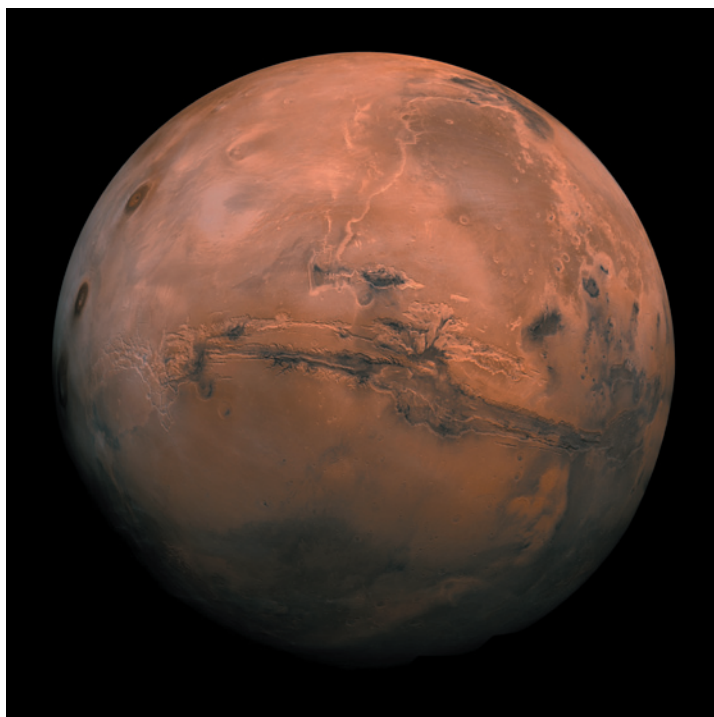
Geological evidence suggests that water once flowed on Mars.

across the Martian surface. (Today the planet is essentially bone-dry, although water may occasionally seep across the surface or just beneath it.)

Mars experts struggled last week to reconcile their detailed geological maps with their inability to explain how water could exist for extended periods. If, billions of years ago, the Martian atmosphere had consisted mostly of carbon dioxide as scientists suspect, it would not have kept the planet warm enough for liquid water to exist for long periods of time. Mars instead might have been much like Antarctica today, with high-elevation areas that remain frozen even at the height of summer. "This is completely different than what we thought about before," says James Head, a planetary geologist at Brown University in Providence, Rhode Island.

Some modellers are still trying to figure out a way to make a warm and wet Mars work. Planetary scientist Ramses Ramirez, of Pennsylvania State University in University Park, proposes that volcanoes belched out enough hydrogen to help heat the atmosphere and allow liquid water to flow (R. M. Ramirez *et al. Nature Geosci.* **7,** 59–63; 2014). Others invoke a temporary warming brought about by sulphur spewed from ancient volcanoes — but even

that would not have kept the planet toasty enough for liquid water year-round, says Laura Kerber of the Pierre and Marie Curie University of Paris.

Fixing the disconnect could take time — and new technology. Planetary geologist Bethany Ehlmann of the California Institute of Technology in Pasadena says that the next generation of instruments to explore Mars should be able to probe rocks in extreme close-up. Her ideal gadget would be a high-resolution camera combined with a spectrometer that could gather information on minerals in the rock in unprecedented detail. Such close-up analysis could reveal more-precise information to help to unravel the history of a particular rock, how it formed and what that might say about past environments, she says.

Also on Ehlmann's wish list: a way for future rovers to measure the precise ages of rocks on the Martian surface. "The more age-dating, the better," she says. "We have to get at these questions of timing." For now, almost all dates for Martian rocks are relative, obtained by looking at how extensively rock formations have been beaten up by impacts — researchers assume that the more pummelled ones have been around longer. Chemical analysis could provide a much more precise age, allowing scientists to tease out the sequence of events in Martian history.
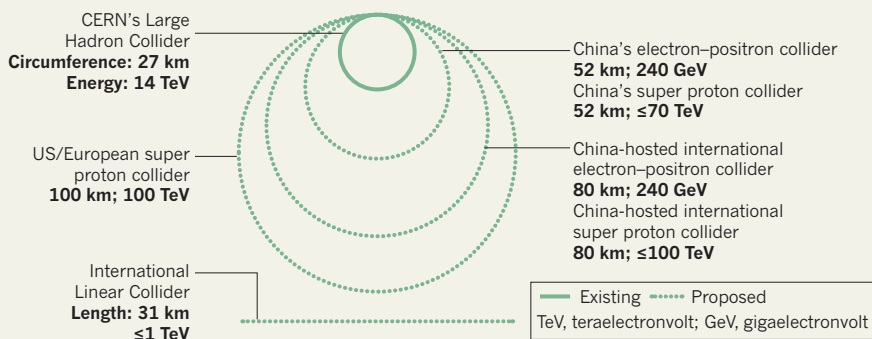
Neither of Ehlmann's goals will be addressed by the next few spacecraft going to Mars: a NASA geophysical probe launching in 2016, and a European orbiter in 2016 and rover in 2018. In the next few weeks, NASA is expected to announce what instruments will be aboard a rover due to launch in 2020, the first step towards gathering samples to be left on the surface for possible future return to Earth (see *Nature* **509,** 272; 2014).

Mars researchers should not be discouraged by the slow progress so far, says Jean-Pierre Bibring, a planetary scientist at the University of Paris-South. The very fact that there are so many data to argue about is a huge step forward, he told the meeting: "It's a dramatic revolution we're in now." ∎

NASA/JPL-CALTECH

COLLISION COURSE

Particle physicists around the world are designing colliders that are much larger in size than the Large Hadron Collider at CERN, Europe's particle-physics laboratory.

CERN's Large Hadron Collider
Circumference: 27 km
Energy: 14 TeV

China's electron–positron collider
52 km; 240 GeV
China's super proton collider
52 km; ≤70 TeV

China-hosted international electron–positron collider
80 km; 240 GeV
China-hosted international super proton collider
80 km; ≤100 TeV

US/European super proton collider
100 km; 100 TeV

International Linear Collider
Length: 31 km
≤1 TeV

——— Existing   ·········· Proposed
TeV, teraelectronvolt; GeV, gigaelectronvolt

their ability to host a 52-kilometre machine, says Ian Shipsey an experimental physicist at the University of Oxford, UK. But after several successes in collider and neutrino experiments, including showing in 2012 how neutrinos change from one form to another, China now has "the confidence, for the first time, to propose an ambitious new machine", says Shipsey.

The Chinese government is yet to agree on any funding, but growing economic confidence in the country has led its scientists to believe that the political climate is ripe, says Nick Walker, an accelerator physicist at DESY, Germany's high-energy physics laboratory in Hamburg. Although some technical issues remain, such as keeping down the power demands of an energy-hungry ring, none are major, he adds.

But China is still a long way from collider dominance. Its main weakness is that its high-energy-physics community is small, says Guido Tonelli, a particle physicist and former head of one of the two major experiments at CERN. If China is to eventually host a super collider, the project will have to be international, he adds. "Nobody would be able to do that alone."

Wang says that China would welcome international funding contributions for both projects, and that if there is a lot of support the ring size could be expanded to 80 kilometres, increasing the scientific scope (see 'Collision course'). But he adds that the country will not wait for collaborators before pressing ahead. The next two years will be spent sketching out a design and establishing what technical difficulties need to be ironed out. Detailed design, budget and location plans will follow, and construction could begin in as little as five years, adds Wang.

But because, realistically, only one super collider will ever be built, says Shipsey, "the world will have to work together to locate it in the best place" for it to happen as quickly as possible. He believes that, in the next five years, the Chinese plans will produce positive competition between China, the United States and Europe, maximizing the chances of a single contender emerging.

"The world will have to work together to locate it in the best place."

There is a final complication. Plans are well under way for an International Linear Collider (ILC), an electron–positron linear accelerator that could operate at much higher energies than China's proposed 52-kilometre electron–positron ring. Physicists are strongly behind the project, but it is yet to secure funding or a host country, says Brian Foster, who leads the European ILC design team. He fears that the Chinese plans could dampen support for the project.

Japan has shown a strong interest in hosting the ILC, but has not "got a stranglehold" on the project, says Foster. He suggests that China could step in, and argues that because the ILC has a wider energy range than a ring collider, the linear accelerator could do more than study the Higgs: it could explore other poorly understood particles, such as the top quark, and any other phenomena that the LHC might discover. ∎
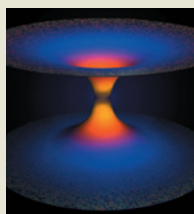
MORE ONLINE

A. CORICHI/J.P. RUIZ

THE
FUSION UPSTARTS

FUELLED BY VENTURE CAPITAL AND A LOT OF HOPE, ALTERNATIVE FUSION TECHNOLOGIES ARE HEATING UP.

BY M. MITCHELL WALDROP

To reach one of the world's most secretive nuclear-fusion companies, visitors must wind their way through a suburban office park at the foot of the Santa Ana Mountains, just east of Irvine, California, until they pull up outside the large but unmarked headquarters of Tri Alpha Energy.

This is as close as any outsider can get without signing a non-disclosure agreement; Tri Alpha protects its trade secrets so tightly that it does not even have a website. But the fragments of information that have filtered out make it clear that the building houses one of the largest fusion experiments now operating in the United States. It is also one of the most unconventional. Instead of using the doughnut-shaped 'tokamak' reactor that has dominated fusion-energy research for more than 40 years, Tri Alpha is testing a linear reactor that it claims will be smaller, simpler and cheaper — and will lead to commercial fusion power in little more than a decade, far ahead of the 30 to 50 years often quoted for tokamaks.

That sounds particularly appealing at a time when the world's leading fusion project, a giant tokamak named ITER, is mired in delays and cost overruns. The facility, being built in Cadarache, France, is expected to be the first fusion reactor capable of generating an excess of energy from a sustained burn of its plasma fuel. But it looks set to cost as much as US$50 billion — about 10 times the original estimate — and will not begin its first fuelled experiments before 2027, 11 years behind schedule.

With ITER consuming the lion's share of the US fusion-energy budget, fans of alternative approaches have scant government support. But growing impatience with the tokamak technology has spurred the Tri Alpha team and many other physicists in the United States and Canada to pursue different options. Over the past decade and a half, these mavericks have launched at least half a dozen companies to pursue alternative designs for fusion reactors. Some are reporting encouraging results, not to mention attracting sizeable investments. Tri Alpha itself has raised $150 million from the likes of Microsoft co-founder Paul Allen and the Russian government's venture-capital firm, Rusnano.

But that success is bringing increased scrutiny of their bold promises. Tri Alpha "has got very tough problems to overcome as it starts scaling up to reactor size", says Jeffrey Freidberg, a nuclear physicist at the Massachusetts Institute of Technology (MIT) in Cambridge. For example, the company must prove that it can achieve the billion-kelvin temperatures needed to burn the exotic fuel it wants to use, and must demonstrate a practical way to convert the energy output into electricity. Similar questions could be raised about any of the other upstarts, says Stephen Dean, who heads Fusion Power Associates, an advocacy group in Gaithersburg, Maryland. "I don't think you can honestly say that any of these things are at the stage where fusion can be demonstrated quickly," he says.

Will alternative fusion companies be able to sustain their momentum and justify their founders' optimism? Or will they fizzle like so many fusion dreams before them?

FOLLOW THE SUN

In principle, building a fusion reactor is just a matter of imitating the Sun. Take the appropriate isotopes of hydrogen or other light elements, add heat to strip the electrons from the nuclei and form an ionized plasma, then compress that plasma and hold it together for a while,

General Fusion's reactor would use massive pistons to crush fuel in a spinning vortex of liquid lead.
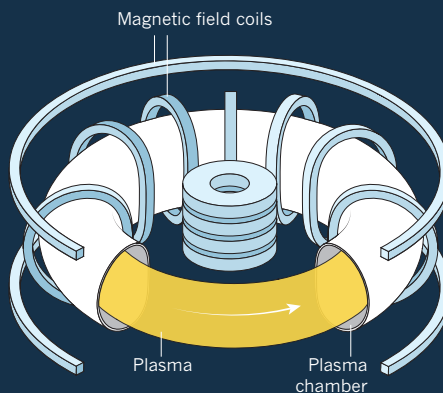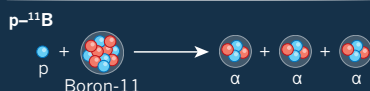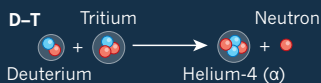
TRAPPING FUSION FIRE

When a superhot, ionized plasma is trapped in a magnetic field, it will fight to escape. Reactors are designed to keep it confined for long enough for the nuclei to fuse and produce energy.


A CHOICE OF FUELS
Many light isotopes will fuse to release energy. A deuterium–tritium mix ignites at the lowest temperature, roughly 100 million kelvin, but produces neutrons that make the reactor radioactive. Other fuels avoid that, but ignite at much higher temperatures.

D–T
Deuterium + Tritium → Helium-4 (α) + Neutron

D–D
D + D → Helium-3 + n

D–³He
D + ³He → α + Proton

p–¹¹B
p + Boron-11 → α + α + α

TOKAMAK
(ITER AND MANY OTHERS)
Multiple coils produce magnetic fields that hold the plasma in the chamber. A coil through the centre drives a current through the plasma to keep it hot.

Magnetic field coils
Plasma
Plasma chamber
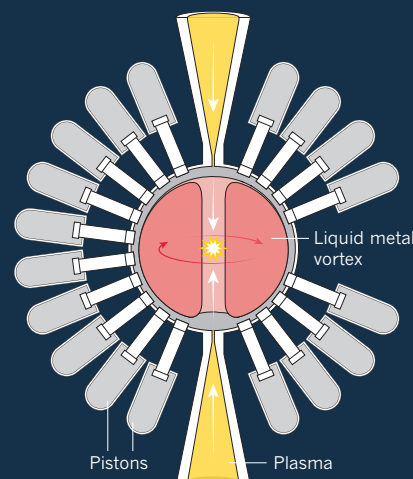
MAGNETIZED TARGET REACTOR
(GENERAL FUSION)
Magnetized rings of plasma are injected into a vortex of liquid metal. Pistons punch the metal inwards, compressing the plasma to ignite fusion.

Liquid metal vortex
Pistons
Plasma
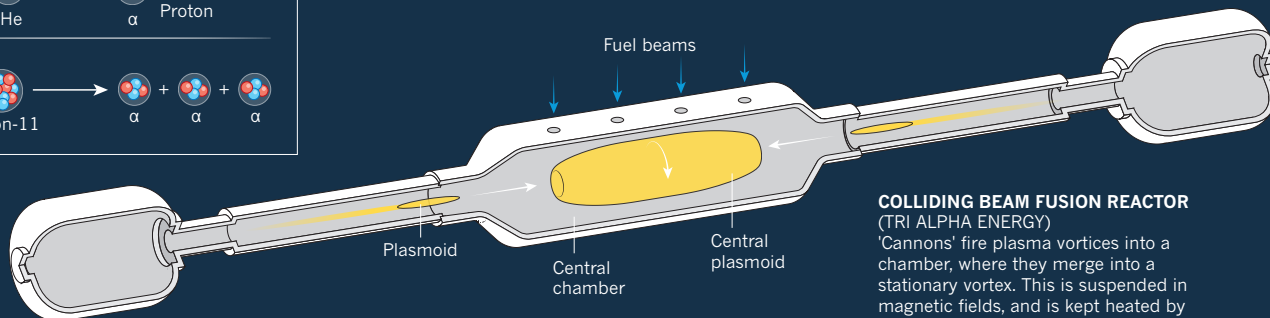
COLLIDING BEAM FUSION REACTOR
(TRI ALPHA ENERGY)
'Cannons' fire plasma vortices into a chamber, where they merge into a stationary vortex. This is suspended in magnetic fields, and is kept heated by beams of fresh fuel.

Fuel beams
Plasmoid
Central chamber
Central plasmoid

allowing the nuclei to fuse and convert a portion of their mass into energy. But in practice, trying to mimic a star leads to horrendous engineering problems: for example, hot plasma trapped in a magnetic field tends to twist and turn like an enraged snake struggling to escape.

Fusion researchers have long favoured tokamaks as the best way to contain this plasma beast. Developed by Soviet physicists in the 1950s and announced to the West a decade later, the reactors achieved plasma densities, temperatures and confinement times much higher than any machine before them. And as physicists refined the design, they improved the way that tokamaks controlled high-energy plasma.

But from the beginning, many physicists have wondered whether tokamaks could ever be scaled up to achieve commercial power output. They are dauntingly complex, for starters. The toroidal chamber has to be wound with multiple sets of electromagnetic coils to shape the magnetic field that confines the plasma. And more coils run through the doughnut hole to drive a powerful electric current through the plasma (see 'Trapping fusion fire').

Then there is the fuel, a mixture of the hydrogen isotopes deuterium (D) and tritium (T). D–T is widely regarded as the only sane choice for a power reactor because it ignites at a lower temperature than any other combination — only about 100 million kelvin — and releases much more energy. But 80% of that energy emerges from the reaction in the form of speeding neutrons, which would wreak havoc on the walls of a power reactor, leaving them highly radioactive. To generate electricity, the neutrons' energy would have to be used to heat water in a conventional

steam turbine — a process that is only 30–40% efficient.

Cost, complexity and slow progress have also dogged inertial-confinement fusion, the most prominent alternative to the tokamaks' magnetic confinement. This approach, in which frozen fuel pellets are imploded by high-powered laser beams, has also received a lot of government funding. But despite decades of effort on inertial confinement, initiatives such as the National Ignition Facility at Lawrence Livermore National Laboratory in Livermore, California, are still struggling to deliver on their fusion-power promises (see *Nature* **491,** 159; 2012).

RADICAL DEPARTURE

Such concerns have sparked some enthusiasm for the stellarator: a toroidal device that simplifies certain aspects of the tokamak but requires even more complex magnets. But most mainstream plasma physicists have simply left the practical engineering issues for later, assuming that fixes will emerge after the plasma physics has been worked out. The fusion mavericks are among the minority who argue that a more radical solution is needed: first get the engineering right, by designing a simple, cheap reactor that power companies might actually want to buy, and then try to make the plasmas behave.

One of those upstarts is Norman Rostoker, a physicist at the University of California, Irvine, who co-founded Tri Alpha in 1998 at the age of 72. He and his colleagues proposed ditching D–T fuel in favour of fusing protons with boron-11, a stable isotope that comprises about 80% of natural boron. Igniting this p–¹¹B fuel would require temperatures of about a billion kelvin, almost 100 times as hot as the core of the Sun. And the energy created in each fusion event would be only about half that released by

D–T. But the reaction products would be practically free of troublesome neutrons: the fusion would generate just three energetic helium nuclei, also known as α-particles. These are charged, so they could be guided by magnetic fields into an 'inverse cyclotron' device that would convert their energy into an ordinary electric current with around 90% efficiency.

Burning a billion-kelvin p–^{11}B plasma in a tokamak was out of the question, not least because unfeasibly large magnetic fields would be needed to confine it. So Rostoker and his colleagues designed a linear reactor that looks like two cannons pointed barrel to barrel. Each cannon would fire rings of plasma called plasmoids that are known to be remarkably stable: the flow of ions in the plasma would generate a magnetic field, which in turn would keep the plasma confined. "It's the most ideal configuration you could imagine," says Alan Hoffman, a plasma physicist at the University of Washington in Seattle.

To start the reactor, each cannon would fire a plasmoid into a central chamber, where the two would merge into a larger, free-floating plasmoid that would survive for as long as it could be fed with additional fuel. The α-particles emerging from the reaction would be guided back through the cannons by another magnetic field, and captured in the energy converter.

By the time the team published this concept[1] in 1997, it was becoming clear that the US energy department was not going to fund development of the machine, preferring instead to focus on tokamaks, which seemed to be a safer bet. "The big experiments have been funded for decades, so there's little chance they won't meet their milestones," says John Slough,

WILL FUSION COMPANIES BE ABLE TO SUSTAIN THEIR MOMENTUM — OR WILL THEY FIZZLE?

a plasma physicist at the University of Washington. "If they start funding these alternatives, all the uncertainties come back."

So Rostoker and his colleagues decided to take advantage of the United States' robust culture of high-tech startups and venture-capital funding. They formed a company, naming it Tri Alpha after the output of the p–^{11}B reaction, and went on to raise enough investment to employ more than 100 people.

Dean suspects that the start-up mindset may explain why Tri Alpha is so secretive. "It's part of the mystique of being a venture-capital-funded company: develop your ideas before anyone else can see them," he says. But over the past five years or so, the company has started to let its employees publish results and present at conferences. With its current test machine, a 10-metre device called the C-2, Tri Alpha has shown that the colliding plasmoids merge as expected[2], and that the fireball can sustain itself for up to 4 milliseconds — impressively long by plasma-physics standards — as long as fuel beams are being injected[3]. Last year, Tri Alpha researcher Houyang Guo announced at a plasma conference in Fort Worth, Texas, that the burn duration had increased to 5 milliseconds. The company is now looking for cash to build a larger machine.

"As a science programme, it's been highly successful," says Hoffman, who reviewed the work for Allen when the billionaire was deciding whether to invest. "But it's not p–^{11}B." So far, he says, Tri Alpha has run its C-2 only with deuterium, and it is a long way from achieving the extreme plasma conditions needed to burn its ultimate fuel.

Nor has Tri Alpha demonstrated direct conversion of α-particles to electricity. "I haven't seen any schemes that would actually work in practice," says Martin Greenwald, an MIT physicist and former chair of the energy department's fusion-energy advisory committee. Indeed, Tri Alpha is planning that its first-generation power reactor would use a more conventional steam-turbine system. Other fusion entrepreneurs will have to tackle similar challenges, but that has not deterred them. Slough is chief scientific officer at Helion Energy in Redmond, Washington, which is developing a linear colliding-beam reactor that would be small enough

to be carried on the back of a large truck. The Helion reactor will fire a steady stream of plasmoids from each side into a chamber, where the fuel is crushed by magnetic fields until fusion begins. Within one second, the fusion products are channelled away just as the next pair of plasmoids hurtles in. "The analogy we like to make is to a diesel engine," says the company's chief executive, David Kirtley. "On each stroke you inject the fuel, compress it with the piston it until it ignites without needing a spark, and the explosion pushes back on the piston."

Helion has demonstrated the concept[4] in a D–D reactor with plasmoids that fire once every three minutes, and it is now seeking $15 million in private financing over the next five years to develop a full-scale machine that could use D–T fuel to reach the break-even point, when it generates as much energy as it takes to run. The company hopes that its reactor could eventually reach the hotter conditions needed to fuse deuterium with helium-3, another combination that produces only α-particles and protons, with no neutron by-products.

Kirtley is optimistic about the money. "There is a giant market need for low-cost, safe, clean power," he says. "So we're seeing a big push in the private investment community to fund alternative ways to generate it." And if the fund-raising is successful, says Kirtley, "our plan is to have our pilot power plant come online in six years."

IN A SPIN

Other alternative concepts stick with D–T fuel, but confine it in different ways. In Burnaby, Canada, researchers at General Fusion have designed a reactor in which a plasmoid of D–T will be injected into a spinning vortex of liquid lead, which will then be crushed inwards by a forest of pistons. If this compression happens within a few microseconds, the plasma will implode to create fusion conditions[5]. One advantage of this design is that the liquid lead does not degrade when it gets blasted by neutrons, says Michel Laberge, who founded General Fusion in 2002.

General Fusion has demonstrated the idea with a small-scale device, using pistons driven by explosives, and has raised about $50 million from venture capitalists and the Canadian government. If the company can win another $25 million or so, Laberge says, it will build a beefier implosion system that can compress the plasma to the levels needed for fusion — perhaps within the next two years.

Despite such optimism, Dean estimates that it will be at least a decade, maybe a lot longer, before any alternative fusion company produces a working power plant. There is simply too much new technology to be demonstrated, he says. "I think these things are well motivated, and should be supported — but I don't think we're on the verge of a breakthrough."

It is not clear how much of that support will come from the US energy department in the foreseeable future. The department's fusion-energy programme has provided a modicum of cash for Helion, as well as for some small-scale academic work on alternative reactors. And its long-shot funding agency, the Advanced Research Projects Agency—Energy, has expressed interest in some of the alterative concepts, to the extent of holding a workshop on them last year. The fusion-energy advisory committee is preparing a ten-year research plan, due by the start of next year, that could conceivably lead to more backing for the upstarts. But funds are tight, and ITER continues to be a huge financial drain.

For now, the big money will probably have to come from the private sector. And despite the many technical hurdles, investors seem willing to take a chance.

"People are starting to think, 'Hey, maybe there are other ways of doing this!'" says Slough. "Maybe it's worth a few million to find out." ■ SEE EDITORIAL P.383

M. Mitchell Waldrop *is a features editor for* Nature *in Washington DC.*

1. Rostoker, N., Binderbauer, M. W. & Monkhorst, H. J. *Science* **278,** 1419–1422 (1997).
2. Binderbauer, M. W. *et al. Phys. Rev. Lett.* **105,** 045003 (2010).
3. Tuszewski, M. *et al. Phys. Rev. Lett.* **108,** 255008 (2012).
4. Slough, J., Votroubek, G. & Pihl, C. *Nucl. Fusion* **51,** 053008 (2011).
5. Laberge, M. *J. Fusion Energy* **27,** 65–68 (2008).

Hello, Governor

When California's governor enlisted the aid of two palaeoecologists, their careers took an unusual turn.

BY VIRGINIA GEWIN

Anthony Barnosky first thought the e-mail was a joke. As an expert in species extinctions who had just sounded the alarm over looming environmental crises, he had grown used to bizarre messages filling his inbox. There was the creepy e-mail enjoining him to "waste himself" to reduce carbon emissions, a plug for carbon-storing "biorocks" and a note encouraging him to explore "yogic flying" to help humanity. So Barnosky was understandably suspicious of the one-line request to contact California governor Jerry Brown. He wearily left a voice message at the number listed, apologizing if he had been spoofed. Out on a run later that day, his phone rang. Brown was on the other end.

It was June 2012 and Barnosky, a palaeo-ecologist at the University of California, Berkeley, had days earlier published a head-line-grabbing *Nature* paper claiming that Earth faced a 'planetary-scale tipping point' because of human-caused climate disruptions, species extinctions, ecosystem loss, pollution and population growth[1]. Among his co-authors was his wife, Elizabeth Hadly, also a palaeoecologist, at Stanford University in California.

The governor had seen the media coverage and had questions about the science but was particularly interested in the level of agreement within the scientific community. Barnosky says that the gist of the conversation came down to one question: "Why aren't you guys shouting this from the rooftops?"

"We thought we were," recalls Hadly.

Brown called a few more times before he made an unusual request: could Barnosky and Hadly translate the science into a format that

he could use in political circles — a consensus statement? They agreed, and with Brown's help created a rallying cry of a report that has received more than 3,300 signatures of endorsement, the majority from researchers.

The document armed Brown with a powerful scientific rationale for the need to combat climate change, and he has handed the report to dozens of politicians, including US President Barack Obama and Chinese President Xi Jinping. Soon after, California entered into an agreement with China to cooperate on developing green technology and reducing greenhouse-gas emissions. And language from the report became part of a climate pact between California, Oregon, Washington and British Columbia, to base energy and environmental decisions on findings presented in the consensus statement as well as in the latest report from the Intergovernmental Panel on Climate Change.

Barnosky and Hadly say that working with Brown on the consensus statement altered their careers in ways they could not have imagined. It monopolized their time for much of a year, made them think much more practically about the relevance of their work and forced them to confront head-on the debate over whether scientists should step forward as policy advocates.

"The consensus statement is more valuable than anything else I've done in my career," says Hadly. "We never could have guessed the reach this paper has had."

TRANSLATING THE TEXT

Before the document could help to inform international negotiations, Barnosky and Hadly had to transform the seven-page paper they had written for scientists into a document aimed at world leaders, policy-makers and the public. Already on sabbatical to write a book, Barnosky took the lead in writing the 46-page statement. He and Hadly discussed
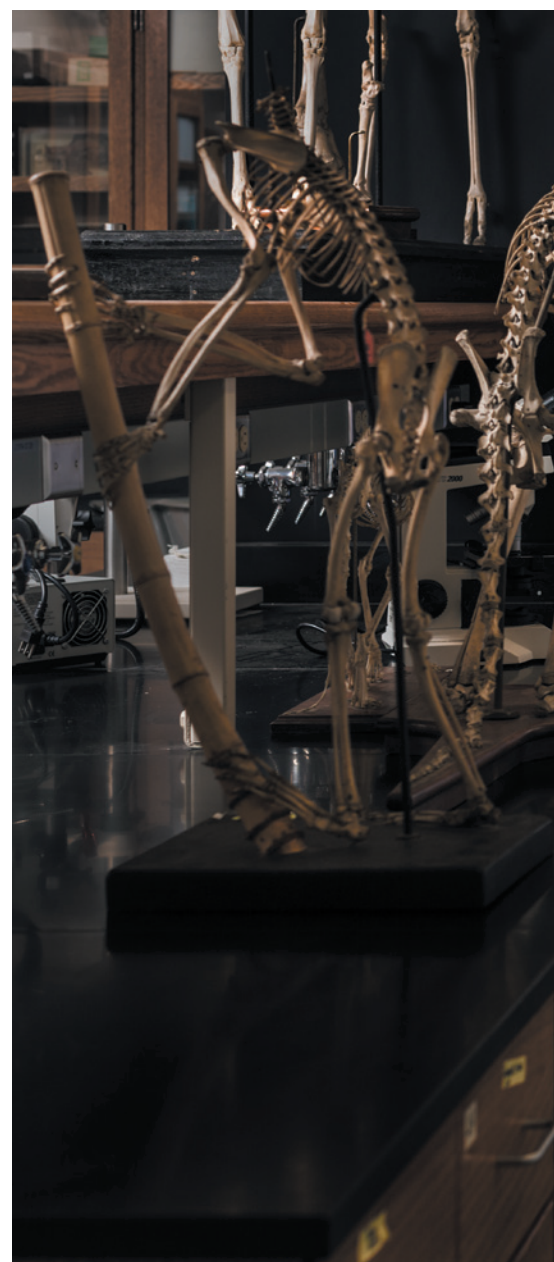
The gist of the conversation came down to one question:

"Why aren't you guys shouting this from the rooftops?"

Anthony Barnosky and Elizabeth Hadly in the laboratory.

the structure while working on it at night and weekends from their home in Palo Alto, California. Barnosky would write draft sections in intensive bursts and send them around to Hadly and the 14 other co-authors. The hardest part, he says, was summarizing the 126 cited studies — without using any scientific jargon. It took 21 iterations to nail down the wording.

This was new territory for them, but they had a guide. "Governor Brown taught us how to do this. He told us what kind of format he needed, not just to understand, himself, but to present to policy-makers," says Hadly. Bulleted points were a must, as was a one-page summary up front. Brown wanted it classic looking, not flashy or cluttered. They went

back and forth on formatting, even where to put the signatures. And the font was key. Brown wanted a simple clear font, Franklin Gothic, with the words 'scientific consensus' highlighted in red. But he confined his advice to style; he had no input on the content.

Once Hadly had put the finishing touches to it, the next step was to solicit signers. She and Barnosky made a list of global-change researchers they knew and sent a mass e-mail to them asking them to sign — and to pass it on to other relevant scientists. Within a month, the report had made its way to 41 countries and garnered 522 signatures. They were pleased with the response, but some people declined to sign because they did not agree with every sentence or, in particular, with the term 'tipping point'.

Although there is little disagreement

that abrupt shifts occur in Earth systems, including climate and the composition of ecosystems, some scientists baulk at the suggestion that there is enough evidence to predict a single tipping point for the whole planet. "I thought it was a great review of the evidence for rapid shifts in ecology, but then it switched to a series of unsupported statements — at best a hypothesis — about how a global tipping point in the biosphere could happen," says Erle Ellis, a landscape ecologist at the University of Maryland, Baltimore County, who was involved in a response to the paper[2].

Yet Ellis understands why the term appeals to politicians. "It's an extraordinarily simple way to look at human-induced global change. It effectively creates a binary Earth; a line drawn in the sand," he says. "Doing
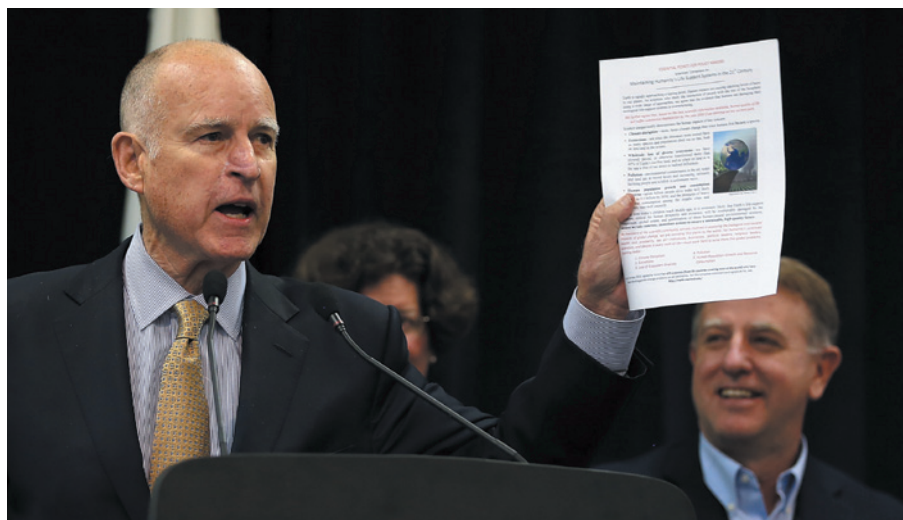
Governor Jerry Brown presents the consensus statement at NASA's Ames Research Center last year.

so gives a false sense of security on the 'safe side' and a false sense it is too late to act on the other." But the concept has power. In fact, game-theory simulations have shown that the kind of coordination needed to solve global environmental problems is much easier to achieve if a tipping point can be predicted with high certainty[3].

So the feature of the paper that Brown, as a lifelong politician, instinctively responded to — the tipping point — was the hardest to sell to scientists. Hadly and Barnosky admit that it is a loaded term. But as researchers who study time periods written in layers of rock, they are used to coming across tipping points. The single scrape of a trowel can reveal, for example, signs of the abrupt extinction of nearly half the world's megafauna some 11,000 years ago. "The world looks different — the chemistry, biology, even the stratigraphy — for a long time after tipping points accumulate and extinctions take place," says Hadly.

To the authors on the consensus statement, the accelerating pace of change on Earth today is sending the planet towards a similar pivot point. They embodied that concept by commissioning Hadly's technician (also an artist), Lily Li, to create a computerized illustration of Earth teetering on the edge of a cliff, held back by a lone, stylized person.

AT THE PRECIPICE

Despite 8 months of work on the consensus statement, neither Hadly nor Barnosky came face to face with Brown until a warm, sunny May morning in 2013, when 400 business, government and civic leaders packed into a conference room at NASA's Ames Research Center in Mountain View, California, for a conference on sustainability technology.

The two scientists were out of their element, with Barnosky in a rarely worn suit and Hadly feeling a bit odd in high heels and a new purple dress. They presented the governor with the report, officially called

Scientific Consensus on Maintaining Humanity's Life Support Systems in the 21st Century. They also gave him a framed version of *Earth on the Cliff*. Brown pointed to the figure holding back the planet and asked who it was. Hadly replied: "It's you."

In his address to the crowd, Brown chastised the media for its anaemic coverage of climate change. He said that a different approach was required to achieve the critical mass needed to create change — something like the consensus statement.

"Governor Brown is a rare politician, as far as his own interest in science and his belief that science can help to persuade the public on climate change," says Susanne Moser, a climate-change communication consultant in Santa Cruz, California, who has worked with Brown on several issues.

As the report's roll-out continued, it was not uncommon for Barnosky and Hadly to receive urgent requests from Brown's office. One day they got a call asking for hard copies of the consensus statement that the governor could take to southern California for a meeting with President Obama and President Xi Jinping. They went to a printing shop, ran off two dozen copies, then Barnosky and his daughter drove the reports to Sacramento, a six-hour round trip.

In the thick of the release, Hadly and Barnosky spent up to half their time working on it. Hadly says that her efforts sparked several conversations with students, who were curious about how to take action without losing respect as a scientist. Conversely, outspoken population biologist Paul Ehrlich from Stanford was sceptical that anything would come of the statement; he had been involved in similar efforts in the past, and gained little political traction.

All along, Hadly and Barnosky have tried to walk a fine line between vigorous communication of the facts and outright advocacy for particular policies. They call their chosen

middle ground "information advocacy", saying that it offers politicians scientifically sound paths on issues but stops short of calling for a particular route. Last autumn, for example, they were asked to add their names to a list of 20 scientists, including Ehrlich and Ken Caldeira, an atmospheric scientist at the Carnegie Institution for Science in Stanford, who were sending an open letter to Brown requesting a ban on shale fracking for oil and gas in California. But they declined, saying that the message was too policy prescriptive. Hadly also turned down an invitation to advocate for research on bringing back extinct species.

"I respect their approach but I don't buy into the notion that prescriptive statements ruin my credibility," says Caldeira. "Everybody has opinions, and it doesn't do any service to science to keep those opinions secret."

Yet Hadly and Barnosky's approach seems to have worked for Brown, the leader of the world's eighth largest economy. Although calls with Brown's staff have slowed to a weekly check-in, the scientists are still working on getting the message out, most recently briefing California's legislative leaders on climate-change impacts and possible mitigation targets, and on their plans for future projects with the governor's office.

They also continue to champion the document outside the United States. Hadly has had the statement translated into other languages and Brown delivered it to political leaders in Norway, Japan, Mexico, Israel and Malaysia. On 4 June, she skyped into an event in Kathmandu during which members of Nepal's parliament signed the consensus statement, and pledged to address climate change as they draft a new constitution. "The thought that a government — particularly one squeezed in between China and India — is crafting a new constitution that hopes to build on these concerns is really powerful," she says.

Barnosky says that it would not have happened without Brown. "You can have all the consensus statements in the world, but what makes them effective is when somebody in a policy-making position actually uses them," he says.

For him and Hadly, the biggest lesson learned is that "a scientist's job isn't over once a paper is published", he says. Or, as Hadly puts it, scientists can reach a point in their careers when they decide, in a world of limited resources and time, to focus on making a difference. ∎

Virginia Gewin *is a freelance writer in Portland, Oregon.*

1. Barnosky, A. D. *et al. Nature* **486,** 52–58 (2012).
2. Brook, B. W., Ellis, E. C., Perring, M. P., Mackay, A. W. & Blomqvist, L. *Trends Ecol. Evol.* **28,** 396–401 (2013).
3. Barrett, S. & Dannenberg, A. *Nature Clim. Change* **4,** 36–39 (2014).

JUSTIN SULLIVAN/GETTY

COMMENT

Fauja Singh, here aged 100, prepares for Britain's Edinburgh marathon in 2011.

JEFF J. MITCHELL/GETTY

Treat ageing

By 2050, the number of people over the age of 80 will triple globally. These demographics could come at great cost to individuals and economies. Two groups describe how research in animals and humans should be refocused to find ways to delay the onset of frailty.

CLINICAL STUDIES
Prepare for human testing

Luigi Fontana, Brian K. Kennedy *and* Valter D. Longo

The problems of old age come as a package. More than 70% of people over 65 have two or more chronic conditions such as arthritis, diabetes, cancer, heart disease and stroke[1]. Studies of diet, genes and drugs indicate that delaying one age-related disease probably staves off others. At least a dozen molecular pathways seem to set the pace of physiological ageing.

Researchers have tweaked these pathways to give rodents long and healthy lives. Restricting calorie intake in mice or introducing mutations in nutrient-sensing pathways can extend lifespans[2] by as much as 50%. And these 'Methuselah mice' are more likely than controls to die without any apparent diseases[3]. Post-mortems reveal that tumours, heart problems, neurodegeneration and metabolic disease are generally reduced or delayed in long-lived mice. In other words, extending lifespan also seems to increase 'healthspan', the time lived without chronic age-related conditions.
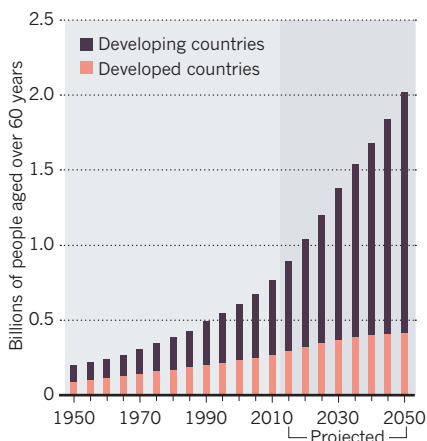
These insights have made hardly a dent in human medicine. Biomedicine takes on conditions one at a time — Alzheimer's disease, say, or heart failure. Rather, it should learn to stall incremental cellular damage and changes that eventually yield several infirmities.

The current tools for extending healthy life — better diets and regular exercise — are effective. But there is room for improvement, especially in personalizing treatments. Molecular insights from animals should be tested in humans to identify interventions to delay ageing and associated conditions. Together, preclinical and clinical researchers must develop meaningful endpoints for human trials.

Longevity pathways identified in model organisms seem to be conserved in humans and can be manipulated in similar ways[2,4]. Genetic surveys of centenarians implicate hormonal and metabolic systems[5]. Long-term calorie restriction in humans induces drastic metabolic and molecular changes ▶

AGE OF THE OLD

By 2050, the number of people aged over 60 years is projected to be five times that in 1950.

gap. Budgets for ageing research are small compared to disease-centred research. The Division of Aging Biology in the US National Institute on Aging receives less than 1% of the National Institutes of Health's budget even though it supports research into the mechanisms underlying most disabilities and chronic diseases. Most grants focus on diseases of specific systems (such as Alzheimer's). Most study sections are not set up to evaluate multidisciplinary research on healthspan. The situation is similar in Europe and Japan.

How should we test interventions that extend healthspan? Human data from dietary restriction and genetic-association studies of healthy ageing could help to channel the most promising pathways identified in preclinical studies. Animal studies should be designed to better mimic human ageing (see 'Track function in ageing animals'). For example, frailty indices are often used in human studies. Comparable indices should be developed for mice.

Suitable endpoints for human trials are needed. Animal work suggests many candidates as potential biomarkers, such as accumulation of molecular damage to DNA, proteins and lipids from oxidative stress. Publicly funded clinical trials — such as one studying potential benefits of aspirin to prevent heart attacks and general decline in the elderly — could also collect crucial samples of blood, muscle and fat for molecular analysis.

Funding agencies should establish committees of translational scientists to review which markers of biological ageing are most consistent between animals and humans, and prioritize the most practical for further assessment. Chosen biomarkers could be evaluated in clinical studies over a broad age range of patients already being treated with drugs that increase lifespan in animal models. Assessments must also be developed for dietary or other interventions that do not involve drugs.

The most important change must be in mindset. Economic incentives in both biomedical research and health care reward treating diseases more than promoting health. The launch of a few anti-ageing biotech companies such as Calico, created last year by Google, is promising. But public money must be invested in extending healthy lifespan by slowing ageing. Otherwise we will founder in a demographic crisis of increased disability and escalating health-care costs.

Luigi Fontana *is professor of medicine and nutrition at Washington University in St. Louis, Missouri, USA, and at Brescia University, Italy.* **Brian K. Kennedy** *is director of the Buck Institute for Research on Aging in Novato, California, USA.* **Valter D. Longo** *is director of the Longevity Institute at the University of Southern California in Los Angeles, USA. e-mail: fontana.luigi@gmail.com*

▶ that resemble those of younger people, notably in inflammatory and nutrient-sensing pathways[6]. Mice engineered to have reduced signalling in these pathways live longer; humans with similar mutations have lower rates of cancer and diabetes[2,7].

Several molecular pathways that increase longevity in animals are affected by approved and experimental drugs[8]. Cancer and organ-rejection drugs such as rapamycin extend lifespan in mice and worms by muting the mTOR pathway, which regulates processes from protein synthesis to cell proliferation and survival. The sirtuin proteins, involved in a similar range of cellular processes, are activated by high concentrations of naturally occurring compounds (such as the resveratrol found in red wine) and extend lifespan in metabolically abnormal obese mice. A plethora of natural and synthetic molecules affect pathways that are shared by ageing, diabetes and metabolic syndrome.

Diet has similar effects. The drugs rapamycin and metformin mimic changes observed in animals fed calorie- and protein-restricted diets. And fasting triggers cellular responses that boost stress resistance, and reduce oxidative damage and inflammation. In rodents, fasting protects against diabetes, cancer, heart disease and neurodegeneration[9]. There are many anti-ageing interventions that could be considered for clinical trials.

IGNORED OPPORTUNITIES

Scientists are not set up to capitalize on these leads to combat the looming ageing crisis (see 'Age of the old'). Clinicians do not realize how much is understood about the molecular mechanisms of ageing and its broad effects on diseases. Laboratory researchers are naive about clinical implications of side effects or overdoses. Researchers of all stripes focus too much on easing or reversing the progression of diseases. Gerontologists should work with people who are middle-aged or even younger.

The problem is calcified by the funding

PRECLINICAL STUDIES
Track function in ageing animals

Douglas Seals *and* Simon Melov

Much is known about how humans age. Over the past decades, gerontologists and geriatricians have developed batteries of tests to assess declines in physiological function as people age. For example, the ability of insulin to clear glucose from blood tracks the risk of diabetes; stiffer arteries indicate a future of hypertension and cognitive impairments, and reduced bone density increases risk of fracture.

Rigorous studies have compared groups of young and older adults or followed the same individuals over time to establish many function-based markers of ageing. Indeed, life expectancy and the likelihood of disability for people over 65 years can be predicted from reductions in grip strength, average walking speed, or the ability to sit and rise from the floor[10,11].

These assessments provide ways to measure whether interventions involving drugs or lifestyle regimens protect the health of older adults. Regulatory authorities in the United States and the European Union are considering using a combination of these kinds of physical-performance tests to evaluate effectiveness in clinical trials of drugs for age-related disorders such as the muscle-wasting condition known as sarcopenia.

At the same time, basic researchers have investigated molecular mechanisms of ageing in several animal models. Genetic and pharmacological manipulations can increase lifespan in worms, flies and rodents[12–14]. There is some evidence that these manipulations might increase healthy lifespan as well (see 'Prepare for human testing'). And centenarians and groups that practise strict healthy lifestyles (such as Seventh-day Adventists, many of whom follow a mostly vegetarian diet) frequently enjoy good health until near the end of life. The rate of ageing, it seems, is malleable, not fixed.

To determine how best to preserve function with ageing in humans, researchers must create common ground between animal studies and clinical assessments. Several functions known to decline with age in humans can also be assessed in rodents (see 'Missing links'). But many physiological tests that have been established in humans either do not exist or are not routinely performed on ageing animals. It is time to expand functional assessments in model organisms, especially rodents.

In 2004, the US National Institute on Aging launched a programme to test strategies that extend lifespan in mammals.

Multiple facilities are funded to assess the same parameters in several mouse strains under standardized conditions[15]. Unfortunately, most measurements occur only after the animals die.

Few studies track declines with ageing in living animals: bones thinning, hearts pumping less efficiently, or sluggishness in cognitive tasks. The first thorough paper on evaluating cardiovascular function in ageing mice was published barely five years ago[16]. Similarly, meetings on the basic biology of ageing concentrate on molecular mechanisms and pay scant attention to function.

Animals have evolved physiological redundancies, so manipulating a gene or signalling pathway does not necessarily change the function of an organ or organism. A drug that slows ageing should slow declines in several organ systems, but few laboratories can measure this in animals. One that does is the Healthspan Assessment Laboratory at the Mayo Clinic in Rochester, Minnesota. The lab quantifies muscle strength and cognitive impairment, and can measure body composition, metabolic rate, insulin sensitivity, motor coordination, bone density and exercise capacity. The Buck Institute for Research on Aging in Novato, California, (where S.M. works) has established similar resources. More facilities like this are needed.

ADAPTING TOOLS

Researchers need new methods to characterize function, particularly to track the same animals as they age, and before and after interventions. They need imaging tools and minimally invasive techniques, and those already established should be included in more experiments. Highest priority should be given to the assessment of motor and cognitive functions[17] most likely to be recognized by drug regulators and industry. And, as in human studies, function should be characterized in several organs and tissues. Variability among animals, which increases with ageing, must also be accounted for.

Even studies that do assess function in animals often do so under conditions with little relevance to human ageing. Experiments on whether a drug improves cardiac function typically use young mice whose hearts have been purposely injured, rather than old mice with naturally aged hearts. To study the effects of obesity on the diseases of ageing, mice are fed a diet of up to 70% fat, with high proportions of trans and saturated fats; such diets bear little resemblance to typical human diets. Effects of inflammation are induced by molecules such as lipopolysaccharide at concentrations that would never be observed in normal physiology. Damage from oxidative stress, which accumulates over the years in humans, is rapidly induced with toxic chemicals such as paraquat.

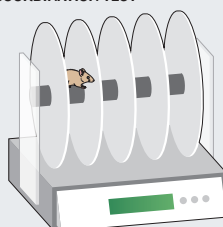Translational studies must be different.

MISSING LINKS

Several measures of how body function in humans changes with age can be adapted to rodents. Other tests not shown here include assessments of heart function (echocardiography), cognition (behavioural tests), aerobic-exercise capacity and metabolic rate.
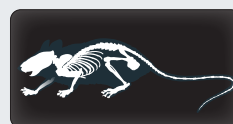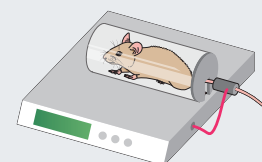
GRIP-STRENGTH TEST
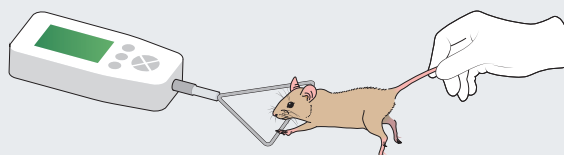
COORDINATION TEST   BODY SCAN   BLOOD-PRESSURE TEST

They should measure parameters such as blood pressure, metabolic rate, cardiac output and kidney filtration at animals' normal resting states and in response to common physiological challenges, such as running on an exercise wheel or completing a maze.

In the past decade, the US National Institutes of Health (NIH) has released a set of testing batteries for clinical research (see go.nature.com/invecq). These measures, including tests for balance, memory, strength and perception, can be used across many kinds of studies and are designed to make it easier to combine results from different trials.

Longitudinal studies in humans, such as the NIH-supported Dynamics of Health, Aging, and Body Composition (Health ABC) Study, have collected body scans, health inventories, and physical performance data (such as walking speed) to establish clinically relevant measures of bone density, blood pressure and motor function. These kinds of assessments, adapted for rodents, should be used to determine which of hundreds of pharmacological compounds and nutritional interventions that are shown to slow the effects of ageing in model organisms should be advanced to trials in humans.

Investigators studying ageing in model organisms and humans must collaborate more. At the NIH, for example, basic and clinical research that relates to the biology of ageing are administered in separate programmes. To translate animal work into human treatments, the ageing-research community must rethink how all those involved communicate and interact, and are trained and funded.

The need for such crosstalk in ageing research has been stated frequently, with little effect. The development of standardized functional assessments that extend from model organisms to human populations may finally make it happen. ∎

Douglas Seals *is professor in the Department of Integrative Physiology at the University of Colorado Boulder in Boulder, Colorado, USA.* **Simon Melov** *is director of genomics at the Buck Institute for Research on Aging in Novato, California, USA.*
e-mail: seals@colorado.edu

1. Hung, W. W., Ross, J. S., Boockvar, K. S. & Siu, A. L. *BMC Geriatr.* **11,** 47 (2011).
2. Fontana, L., Partridge, L. & Longo, V. D. *Science* **328,** 321–326 (2010).
3. Ikeno, Y. *et al. Pathobiol. Aging Age Relat. Dis.* **3,** 20833 (2013).
4. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. *Cell* **153,** 1194–1217 (2013).
5. Barzilai, N. *et al. J. Clin. Endocrinol. Metab.* **95,** 4493–4500 (2010).
6. Mercken, E. M. *et al. Aging Cell* **12,** 645–651 (2013).
7. Guevara-Aguirre, J. *et al. Sci. Transl. Med.* **3,** 70ra13 (2011).
8. Kennedy, B. K. & Pennypacker, J. K. *Transl. Res.* **163,** 456–465 (2013).
9. Longo, V. D. & Mattson, M. P. *Cell Metab.* **19,** 181–192 (2014).
10. de Brito, L. B. *et al. Eur. J. Prev. Cardiol.* **21,** 892–898 (2012).
11. Studenski, S. *et al. J. Am. Med. Assoc.* **305,** 50–58 (2011).
12. Johnson, T. E. *Exper. Gerontol.* **48,** 640–643 (2013).
13. Kenyon, C. J. *Nature* **464,** 504–512 (2010).
14. Harrison, D. E. *et al. Nature* **460,** 392–395 (2009).
15. Nadon N. L. *et al. Age (Dordr.)* **30,** 187–199 (2008).
16. Dai, D.-F. *et al. Circulation* **119,** 2789–2797 (2009).
17. Justice, J. N. *et al. Age (Dordr.)* **36,** 583–595 (2014).

CORRECTION
The Comment 'Polio eradication hinges on child health in Pakistan' (*Nature* **511,** 285–287; 2014) accidentally referred to Abbottabad as a village instead of a city.

SCIENCE VISUALIZATION

Branching arguments

Kevin Padian considers two books on tree diagrams and what they can represent.

When we represent the relationships of organisms, our thoughts turn naturally to trees. We draw genealogical family trees, and often think of great apes as part of our extended family. (Some of them even live in trees.) Trees' progressively smaller branching lends itself as a visual metaphor for the diversification of life.

This is problematic: diversification has not always meant evolution as we think of it today. Nineteenth-century authors from geologist Charles Lyell to naturalist Richard Owen plotted the diversification of fossil organisms through layers of the geological column, but most did not regard it as a transmutation of species. Instead, it was seen as a record of life's history. The processes producing that pattern were not accepted until Charles Darwin published *On the Origin of Species* in 1859.

Darwin convinced people that all living beings had one or a few common ancestors, so treelike diagrams became, for the first time, literal genealogical representations of those relationships. However, Victorian authors who used them were not always depicting evolutionary relationships — they could just be showing affinities. All sorts of connecting diagrams can unite creatures that have four digits, are insectivores or live in New Zealand, without suggesting any genealogical connection.

Two books now get to grips with the tree as a pictorial device. Biologist David Archibald's *Aristotle's Ladder, Darwin's Tree* interweaves the history of treelike diagrams with their changing meanings as they came to represent different kinds of statement about relationships. Designer Manuel Lima's *Book of Trees* explores the many possibilities presented by these diagrams in expressing ideas and discoveries, with a strong emphasis on the effectiveness of representational strategies.

Archibald shows how Darwin wrestled
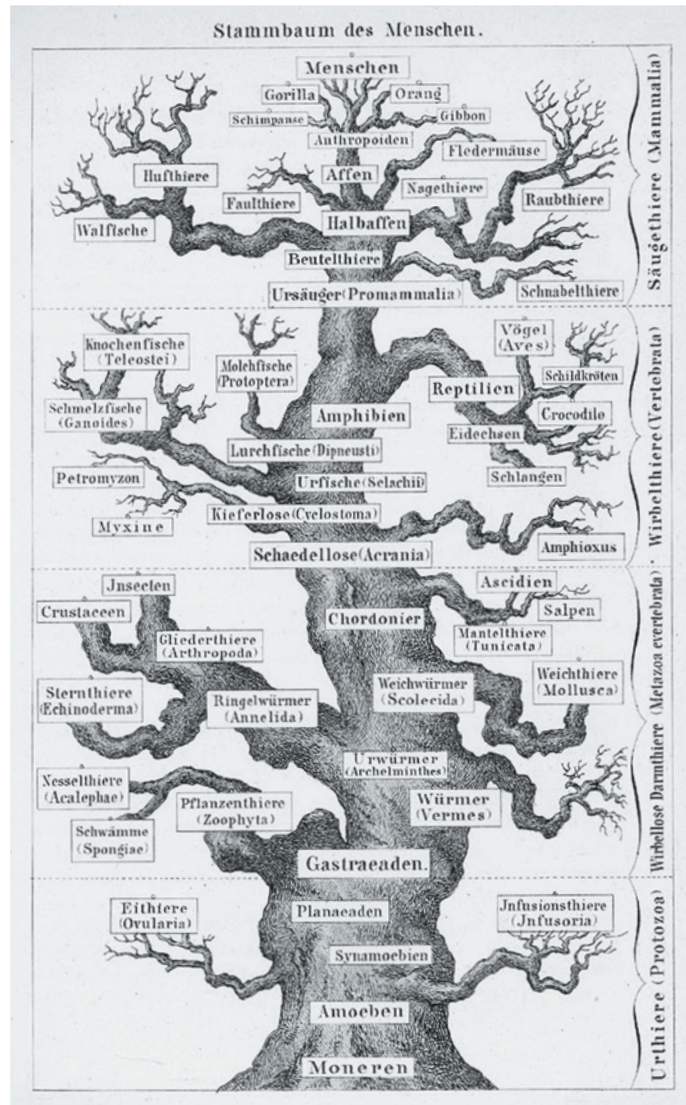
Ernst Haeckel's 1874 'Genealogical Tree of Humans'.

with whether the metaphor of a tree, coral or endlessly dividing seaweed would best represent the genealogies of life. (Corals were attractive because they are alive only at their tips; their nether branches would thus

Aristotle's Ladder, Darwin's Tree: The Evolution of Visual Metaphors in Biological Order
J. DAVID ARCHIBALD
Columbia University Press: 2014.

The Book of Trees: Visualizing Branches of Knowledge
MANUEL LIMA
Princeton Architectural Press: 2014.

represent the vast and largely unknown fossil record.) The book tracks tree iconography back to ancient Rome, where it was used to portray family histories of prominent citizens, and traces the same leafy configurations through early Christian schemes of prophets and saints. How could such a useful device not have come to serve so many conceptual purposes?

Discussing the phenomenon now, Archibald shows that in the modern synthesis of evolution — the neo-Darwinian movement that dominated evolutionary theory from the 1930s — taxonomies were originally relatively free-wheeling. Taxonomists ranked organisms in arbitrary systems according to perceived differences or importances; thus, humans could be separated from (other) apes into a different sub-family, family or even order. Archibald also clarifies cladistics, the current pre-eminent system of systematics, in which genetic proximity is the only criterion for building trees of life (as Darwin thought). In cladistics, newly evolved characteristics (synapomorphies) are prized for their ability to reflect the common ancestry of all the creatures who possess them, compared with those who do not. The treelike cladogram is the ultimate arrangement of groups of creatures according to shared possession of characteristics — transparent in its data, and stringent in its method.

The devil, of course, is in the details of interpreting characteristics in different creatures and choosing the algorithms by which to group them, a process now eased by computer programs. Archibald shows how the rise of methods for extracting and analysing data on similarities in DNA has also revolutionized systematics. These chapters are among the best short treatments of the subject anywhere.

Even in the current hegemony, however, confusion persists. Just as classical treelike diagrams did not always imply relationships,

phylogenetic trees and cladograms convey different ideas. Many students are taught that 'a tree is a cladogram with time attached'. But cladograms are merely diagrams of the distribution of characteristics. You can make cladograms of chocolate bars, hardware or cartoon characters, as long as you clarify how you are identifying the synapomorphies. A phylogenetic tree, by contrast, depicts evolutionary relationships. You can put putative ancestors in a tree's branches (most no longer do so, because we can seldom identify direct ancestors), but not in a cladogram. A cladogram has no real branches — its lines just connect points. You can calibrate a tree against geological time and the dates of fossils; you can make the branches thinner or thicker to depict their diversities; and you can artistically elongate some branches to suggest temporal, adaptive or ecological differences. Trees have few rules, and so they remain an eternal metaphor for biologists, perpetually dangling the tempting apple.

Archibald's book is interdisciplinary, authoritative, well-written and complete, with a deep historiographic appreciation of its many subjects. I wished for a concluding chapter that integrated all the delightful insights in the book, but that is a quibble. It is important in this context to note biologist Theodore Pietsch's complementary work *Trees of Life* (Johns Hopkins University Press, 2012). With its copious illustrations and economical but enlightening text, Pietsch's book is in some ways intermediate between Archibald's and Lima's.

Lima's *Book of Trees* is a catalogue that includes diagrams from biology, philosophy, technology, history and even comic books. They are not all genealogical trees; the book includes various kinds of diverging diagrams, some not even obviously treelike, such as rectangular concept maps. Some compare and even quantify differences in their variables, using their component lengths, thicknesses and directionalities.

As Archibald shows, Darwin used nearly every kind of treelike diagram in his notes and books. Lima's catalogue explains each of Darwin's beautifully reproduced images, but he does not dwell on analysis. For him, the strength of the representations is in their diversity, ability to convey information, and — let's face it — beauty. Clarity of expression is a wonderful thing: the graceful geometric skeletons of points and lines convey more conceptual dimensions than the page can literally express. ∎

Kevin Padian *is in the Department of Integrative Biology and the Museum of Paleontology at the University of California, Berkeley.*
email: kpadian@berkeley.edu

Books in brief

Adventures in the Anthropocene: A Journey to the Heart of the Planet We Made
*Gaia Vince* CHATTO & WINDUS *(2014)*
When *Nature*'s former news editor Gaia Vince set off on a two-year, six-continent trek, her aim was inspired: to explore empirically, amid biodiversity collapse and global climate change, how the planet and its ingenious humans are faring. Digging beneath the scree of statistics, received wisdom, "shock-doctrines and tired slogans", Vince has reached ground truths. Her broader discussions of the biological and Earth science are as cogent as her close reportage on innovators such as the Ladakhi 'glacier-maker' Chewang Norphel.

The Fourth Revolution: How the Infosphere is Reshaping Human Reality
Luciano Floridi OXFORD UNIVERSITY PRESS *(2014)*
Are we 'inforgs' living in 'onlife' — informational organisms inhabiting a blurry realm between real and virtual? Information ethicist Luciano Floridi thinks so, and his searing study of our digital dependency peels to the bone the implications for everything from identity to the environment. We look with new eyes at our transformation into generic online consumers ("Gogol's dead souls, but with wallets"), and our creation of an environment that is dumbed-down enough for smart technologies to excel. Non-alarmist and very, very smart.
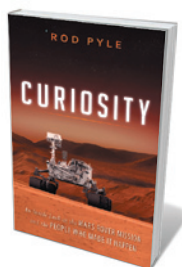
Curiosity: An Inside Look at the Mars Rover Mission and the People Who Made It Happen
*Rod Pyle* PROMETHEUS BOOKS *(2014)*
Ebullient science writer Rod Pyle follows his 2012 *Destination Mars* with more hot reportage of the red planet in this comprehensive study of NASA's Mars Science Laboratory mission and the US$2.5-billion rover Curiosity. Pyle veers from the rover's battery of analytical instruments (one, notably, vaporizes rock) to its findings and operation, as well as the big personalities who make it happen. Gripping behind-the-scenes encounters abound, not least with the dedicated crew that controls the robotic arm remotely.

Music and the Making of Modern Science
Peter Pesic MIT PRESS *(2014)*
Mathematics, astronomy and music were closely allied in classical Greece. That chord has reverberated down the centuries, argues musician and science writer Peter Pesic. The fourteenth-century French philosopher Nicole Oresme, for instance, was an early proponent of heliocentrism who looked at celestial harmony in terms of music and geometry. Centuries later, physicist (and pianist) Max Planck's experiments with harmonium tuning may have inspired his thinking on black-body radiation, Pesic avers: there are analogies between tuning and the behaviour of electromagnetic waves.
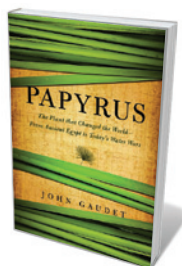
Papyrus: The Plant that Changed the World — From Ancient Egypt to Today's Water Wars
John Gaudet PEGASUS BOOKS *(2014)*
In this swirling anthropological and environmental narrative, ecologist John Gaudet reminds us how papyrus was historically central to swathes of Africa. Gaudet takes us from the continent's ancient marsh-based cultures — in which the reed served as food, a construction material and the earliest form of paper — to today's imperilled wetlands. The reed beds should be reinstated, he argues: they are avian habitats and ideal heavy-metal filters. **Barbara Kiser**

Correspondence

Brain secrets cheap at twice the price

The Apollo space missions cost more than US$100 billion in today's terms, the Large Hadron Collider $10 billion and the development of the Airbus A380 $15 billion — making the projected $4.5 billion for the BRAIN (Brain Research through Advancing Innovative Neurotechnologies) initiative look paltry by comparison (see *Nature* http://doi.org/tmh; 2014).

If we spent that much putting a man on the Moon, smashing particles and building a flying hotel, why not spend at least $10 billion trying to solve the greatest mystery of all?

Anuj K. Rastogi *University of Toronto, Canada.*
anuj.rastogi@mail.utoronto.ca

Overcome hurdles to global databases

We suggest some pointers to guarantee the long-term storage, accessibility and reliability of international research data sets.

Data owners, peer-reviewed journals, research institutes and universities have cited various problems with the feasibility of establishing global databases — including maintenance costs, restrictive data-exchange policies and conflicts of interest.

Research organizations need to start implementing management plans for their own data, as is happening at Wageningen University in the Netherlands. Institutional policies, for example, should mandate that researchers store their data in public repositories.

Setting up distributed databases that are owned by the institutions responsible for the data could also extend data sharing. These institutions would develop their own data policy, and could, for example, pinpoint and resolve any conflicts of interest at an early stage.

Funding agencies and all journals should use their influence to counter restrictive data-sharing policies — for instance, by ensuring that researchers make their data publicly available soon after publication (see, for example, go.nature.com/o5ykhe).
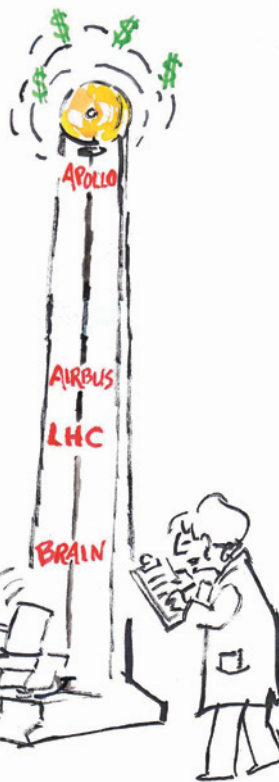
Clear incentives for storing data must replace objections. Collaboration is key to constructing useful databases, both distributed and global.

Venceslas Goudiaby, Pieter A. Zuidema, Godefridus M. J. Mohren *Wageningen University, the Netherlands.*
venceslas.goudiaby@wur.nl

China's land creation project stands firm

We contest the views of Peiyue Li and colleagues on the geological implications of land creation projects in China, particularly for Yan'an city (see *Nature* **510,** 29–31; 2014).

The Yan'an project has created flat ground by levelling hills and filling valleys, and was carefully planned by the local government after thorough feasibility testing by geoengineers, hydrologists, water conservationists and ecologists. The site was consequently able to withstand severe erosion caused by a once-in-a-century flood in July 2013.

Any threat of collapse of soft-soil deposits has been eliminated by repeatedly compacting the relocated soil in layers. This reinforcement strategy has increased soil-particle density from 1.9 to 2.1 grams per cubic centimetre (see go.nature.com/mqg194; in Chinese).

Guided by strict planning and controls, construction work will initially occur in the excavation area, and will not extend to filled regions until the ground stabilizes through natural sedimentation. The soil interfaces between the excavated and filled areas, slope compaction, and drainage of ground and surface water are also being rigorously monitored.

The newly created land will be used for administrative, residential and commercial purposes, so there will be no risk of industrial pollution. Almost half of the project area will be covered in vegetation.

Yan'an's experience of land creation could be useful for extending other mountain cities in China and abroad.

Yansui Liu, Yuheng Li *Beijing Normal University; and Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China.*
liuys@igsnrr.ac.cn

Tactics to stamp out herbicide resistance

The rise of weeds that are resistant to the herbicide glyphosate has led many US farmers to seek permission from the Environmental Protection Agency to use more-dangerous herbicides, such as propazine (see *Nature* **510,** 187; 2014). To deal with this crisis, US agriculture must implement a multi-pronged strategy.

Tactics should include: annual switching between herbicides that have different chemical modes of action; using a variety of methods to remove weed seeds from the soil, including hoeing by hand; and moving to new tillage practices that require a shift from no-till to rotary hoeing and between-row cultivation of the soil to uproot weeds.

US policy-makers need to encourage crop rotation through new subsidies and to impose restrictions on herbicide use. It seems that Europe has dodged a bullet by resisting pressure to sow glyphosate-tolerant crops.

Margaret Mellon *Mellon Associates, Washington DC, USA.*
mardimellon@mac.com

CORRECTION

The Outlook article 'Fat chance' (*Nature* **511,** S14–S15; 2014) incorrectly stated that ice cream is part of a ketogenic diet. In fact, this diet is high in fat but low in carbohydrate so does not include high-sugar foods.

SCHIZOPHRENIA

Genesis of a complex disease

The largest genome–wide analysis of schizophrenia performed so far has identified more than 100 genetic regions that contribute to disease risk, establishing new leads for understanding this form of mental illness. SEE ARTICLE P.421

JONATHAN FLINT & MARCUS MUNAFÒ

A paper[1] published this week in *Nature* marks the culmination of a long debate about the genetic basis of a disorder sometimes considered psychiatry's heartland[2] — schizophrenia. No other psychiatric condition has evoked such diverse opinions. Its biological roots have often been denied and, in the anti-psychiatry movement of the 1970s, there was even outright rejection of its existence. The latest paper, from the Schizophrenia Working Group of the Psychiatric Genomics Consortium, reports an analysis of more than 150,000 people and finds more than 100 genetic regions associated with schizophrenia, laying to rest forever the idea that genetics is not an important cause of the illness.

Dispute over schizophrenia's genetic basis has been ferocious. This is hardly surprising given that Swiss psychiatrist Ernst Rüdin — an early proponent of the argument that the condition is a single-gene disorder — advocated the view that people with mental illnesses should not have children[3], and justified the sterilization and murder of people with schizophrenia. Despite the disease's long and contentious history, by the end of the twentieth century there was consensus that genetic factors were involved[4]. But agreeing that genetics has a part to play is not the same as finding individual genetic regions (loci) that contribute to disease susceptibility. Progress in this arena has been marked by false starts[5,6], and by more than 800 genetic associations[7] of dubious value.

Not unreasonably, many would ask why we should be any more confident that the consortium's geneticists have now got it right. The short answer is that the tests for associations between each gene in the human genome and disease are now mature. The correct criteria for determining significance in these tests are as familiar to human geneticists as their two-times table, and are exhaustively documented in the consortium's paper. Since the first genome-wide analyses for schizophrenia emerged in 2009 (refs 8, 9), the size of the studies has increased, as have the numbers of loci
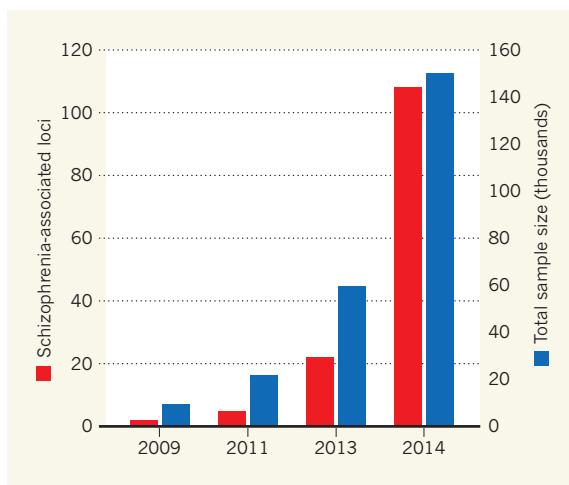

Figure 1 | Progress in identifying schizophrenia-associated genes. Since 2009, genome-wide analyses of the genetic loci associated with schizophrenia (including the current study by the Schizophrenia Working Group of the Psychiatric Genomics Consortium[1], labelled 2014) have yielded increasing numbers of 'hits'[1,9,15,16]. Over this time, there has been a correlation between the number of people tested and the number of susceptibility-associated loci found.

associated with the condition (Fig. 1).

The Psychiatric Genomics Consortium has now confirmed the existence of 108 loci that contribute to disease susceptibility. But even this number is insufficient to entirely explain the genetic causes of schizophrenia. So what does this discovery mean?

First and foremost, it confirms that genetics is a major cause of the illness. The risk variants now identified are common — they contribute in most, if not all, cases. This is a tremendous advance, of the sort that rewrites textbooks. Given the turbulent history of the field, this is a point that deserves emphasis and should be cause for well-earned celebration among those who carried out the work.

Second, we now know something about the probable genes involved. Readers not familiar with the vagaries of genome-wide association studies need to be aware that there is a difference between finding a genetic locus and finding a gene. Usually, it is assumed that the relevant gene is the one that lies nearest to the identified locus, but, as recently reported[10] for a locus involved in obesity, this is not necessarily the case. As a rule of thumb, however,

the closest gene is probably a good place to start.

So what genes are involved? Expectant readers will have to turn to Table 3 of the Supplementary Information (go. nature.com/koxm8i) to find all the details. Those who take the trouble will be rewarded with some delightful insights. For many years, a leading hypothesis has been that schizophrenia arises from abnormal neuronal signalling involving dopamine. This theory was based on evidence[11] indicating that the efficacy of antipsychotic drugs correlates with their ability to block dopamine receptors. It turns out that those who championed the involvement of dopamine-related genes were right (a dopamine-receptor gene, *DRD2*, is associated). However, by no means all dopamine-related genes are involved (the gene encoding catechol-*O*-methyltransferase, an enzyme that metabolizes released dopamine, is not associated).

Having climbed a mountain of genetic analyses, it is perhaps a shame that the authors do not offer much of a view from the peak — no new biological hypotheses are presented. But it is clear that the idea that schizophrenia is merely a disorder of neuronal signalling (either dopamine- or glutamate-related) stands in need of revision. By far the most significant association is a locus on chromosome 6 that includes a region containing genes involved in acquired immunity: the major histocompatibility complex (MHC).

Given that the MHC also contains genes encoding various functional activities, the fact that the most significant association lies in the MHC is only suggestive of its role. However, Figure 2 in the paper contains another intriguing observation. The genetic effects in schizophrenia are enriched in regions outside the MHC that are also involved in acquired immunity. Is schizophrenia in part a disorder of acquired immunity? Surely this idea should start to be taken seriously. It is perhaps ironic that one of the earliest genetic associations reported[12] was with the MHC.

It is possible that environmental exposure explains some of the genetic associations

documented in this report. For instance, DNA-sequence variations in the *CHRNA5-A3-B4* gene cluster are strongly associated with heavy smoking[13]. Smoking is extremely common (more than 80% prevalence) in people with schizophrenia[14], and the consortium's identification of an associated variant in the *CHRNA5-A3-B4* gene cluster might reflect a dose–response relationship between heaviness of smoking and schizophrenia risk, instead of — or as well as — a direct relationship. As ever-larger genome-wide association studies are conducted, we will presumably begin to pick up indirect effects. In other words, such studies may tell us about the environmental as well as the genetic causes of the disease. ■

Jonathan Flint *is at the Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.*
Marcus Munafò *is at the School of Experimental Psychology, University of Bristol, Bristol BS8 1TU, UK.*
e-mail: jf@well.ox.ac.uk

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium. *Nature* **511,** 421–427 (2014).
2. Bebbington, P. & McGuffin, P. (eds) *Schizophrenia: The Major Issues* (Heinemann, 1988).
3. Torrey, E. F. & Yolken, R. H. *Schizophr. Bull.* **36,** 26–32 (2010).
4. Sullivan, P. F., Kendler, K. S. & Neale, M. C. *Arch. Gen. Psychiatry* **60,** 1187–1192 (2003).
5. Brzustowicz, L. M., Hodgkinson, K. A., Chow, E. W. C.,
6. Honer, W. G. & Bassett, A. S. *Science* **288,** 678–682 (2000).
7. Levinson, D. F. *et al. Science* **296,** 739–741 (2002).
8. www.schizophreniaforum.org/res/sczgene
9. The International Schizophrenia Consortium. *Nature* **460,** 748–752 (2009).
10. Stefansson, H. *et al. Nature* **460,** 744–747 (2009).
11. Smemo, S. *et al. Nature* **507,** 371–375 (2014).
12. Seeman, P., Lee, T., Chau-Wong, M. & Wong, K. *Nature* **261,** 717–719 (1976).
13. McGuffin, P. *Psychol. Med.* **9,** 721–728 (1979).
14. Ware, J. J., van den Bree, M. B. M. & Munafò, M. R. *Nicotine Tobacco Res.* **13,** 1167–1175 (2011).
15. Hughes, J. R. *et al. Am. J. Psychiatry* **143,** 993–997 (1986).
16. Ripke, S. *et al. Nature Genet.* **45,** 1150–1159 (2013).
17. Ripke, S. *et al. Nature Genet.* **43,** 969–976 (2011).

EARTH SCIENCE

Rain on the parade

An analysis of landforms in the Bolivian Andes suggests that surface uplift has shaped the climate and landscape. This contrasts with previous work suggesting that climate controls topography and deformation along the mountain range.

ALISON M. ANDERS

Interactions between climate, erosion and deformation of the solid Earth have been hotly debated in recent decades. If climate dictates spatial patterns in erosion, as some researchers conclude, then it may also change mass distribution enough to influence deformation. A pair of papers[1,2] from Whipple and Gasparini, published in *Lithosphere*, now examines relationships between climate, erosion, river-channel steepness and relief in the Bolivian Andes. The authors find unexpectedly little evidence for climatic control of erosion patterns or topography. Instead, they conclude that over the past 12 million years (Myr), the region has uplifted by approximately 3 kilometres.

The tantalizing potential for atmospheric processes to drive deformation of the solid Earth has inspired much research, with results being exuberantly paraded in support of this theory. The central Andean plateau is a compelling region in which to study this issue, because of the correlated spatial variability of its topography, climate and mountain-belt geometry (Fig. 1). Whipple and Gasparini's evidence[1] for recent, large-scale surface uplift comes from low-relief (largely flat) surfaces in the northern Bolivian Andes.

To understand their results, picture the studied region as a series of small, gently sloping plateaus surrounded by steep river canyons on all sides. The authors interpolated the landscape between these plateaus and found
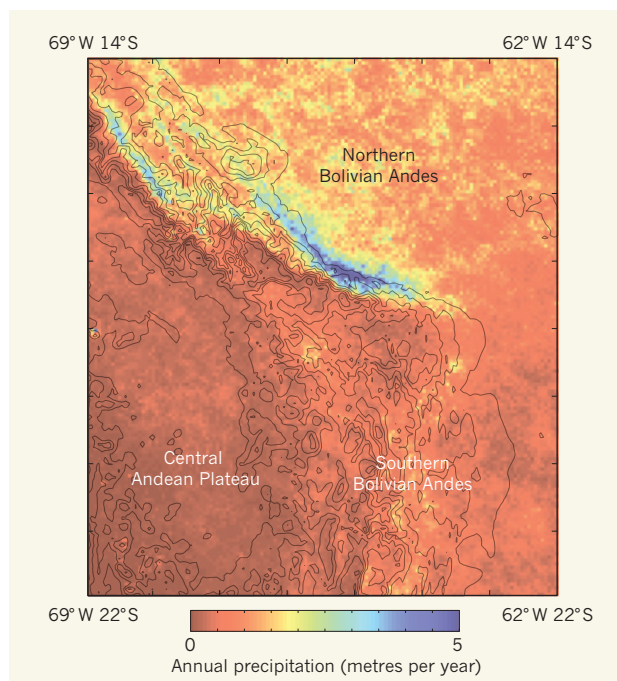

Figure 1 | Precipitation patterns across the Central Andean Plateau. The wet, steep and narrow northern Bolivian Andes strongly contrast with the broad, gently sloped and dry southern Bolivian Andes. Two papers propose[1,2] that the topographic difference between them is driven mainly by changes in tectonic deformation of the mountain belt, and that precipitation varies as a function of topography. Solid lines are topographic contours; the contour interval is 500 metres. Precipitation data shown are from ref. 7.

that they form part of a ramp about 300 km long, gently dipping to the north and grading into sedimentary deposits of approximately 11–9 Myr old. This suggests that the topography of the region about 10 Myr ago was more gently sloping than it is now, and that the land surface has uplifted and tilted since that time.

Flat areas surrounded by incised river valleys are unusual. Typically, there is no difference in elevation at places where tributaries meet; smaller streams maintain steeper slopes to match the pace of erosion of the larger river. However, fluvial hanging valleys — which have floors that are higher than the main channel into which they flow — form in some tectonically active areas where small tributaries intersect with a river that is experiencing a sudden increase in erosion[3]. The stranded small tributaries, separated by waterfalls from the main stream, do not undergo the rapid erosion occurring around them, and are maintained as relicts of the former land surface.

Whipple and Gasparini used river channels on a relict surface to estimate the extent of downcutting of the main river, and assumed that this downcutting was caused by uplift of the surface since its formation. The authors extrapolated trends in channel slopes to reconstruct the channels' positions before the uplift. These reconstructed channels are roughly 2–3 km above current channel bottoms in the incised valleys — in other words, the surface has uplifted by this amount. The timing and magnitude of this surface uplift correspond approximately to estimates of these quantities for the southern Bolivian Andes[4], suggesting that the northern and southern Bolivian Andes evolved in parallel, despite marked contrasts in climate and mountain-belt width for the two regions.

A strong point of these studies is the observation and interpretation

of low-relief surfaces as markers of uplift. The researchers also present a model to explain how the uplift occurs, but this proposes a geometry of faults and folds — specifically, a 'relay ramp' — that has not previously been observed at such a large scale. So could there be an alternative explanation? An example from Australia suggests one possibility.

A mountainous region of Australia called the Adelaide fold-thrust belt reproduces much of the geometry of the Bolivian Andes, including narrowing and widening of the mountain belt with deeper erosion in the narrow portion of the range. The deep exhumation and narrowness of the Adelaide mountain belt results from its impingement on harder rocks in front of the range[5]. The Bolivian mountains studied by Whipple and Gasparini abut a feature called the Chapare Basement High, and the patterns of uplift and erosion observed by the authors are consistent with deformation due to interaction between this region of hard rock and the growing mountain belt in Bolivia.

It is notoriously difficult to discern the impact of climate on landscape topography, because steep and high-relief terrain may result from rock uplift but such terrain also tends to increase precipitation rates. In their second paper, Gasparini and Whipple[2] report a method to isolate climatic effects from tectonic ones. Their approach relies on the fact that precipitation reaches a maximum value midway up steep escarpments. The authors therefore compared the profiles of river channels that drain the dry, upper portion of an escarpment in the northern Bolivian Andes with those from rivers that originate within the wettest part of the topography. The difference in channel steepness reflects the impact of variable precipitation on erosional efficiency. The authors coupled the resulting data with process-based models of river incision, and concluded that climate alone is not responsible for the observed spatial patterns in river steepness and local relief: a gradient in rock uplift is required.

The researchers' analysis of channel profiles, coupled with the evidence of a shared history of surface uplift for the northern and southern Bolivian Andes, supports the view that tectonic deformation controls the topography of the northern Bolivian Andes, whereas climate passively responds. But it is not yet time to cancel the parade. Climate certainly does affect this landscape: the authors' modelling[2] of channel steepness suggests that climate is a secondary influence, and the best-fitting model includes both an uplift gradient and the effects of precipitation. Questions remain about interactions between climate, erosion and tectonics.

This research opens up two new directions for future work. First, we should find out how extreme precipitation gradients can be accommodated without an erosional response. Precipitation must be integrated into river discharge to influence river erosion, which means

that larger rivers are less sensitive to precipitation gradients — the greater the volume of the river, the smaller the effect of adding precipitation. But precipitation gradients also produce variability in ecosystems and soil properties, which in turn influence discharge production. Vegetation and soil remain underexplored links between climate and erosion.

Second, glaciation seems to be a more powerful climatic driver of topographic and tectonic change than are precipitation gradients. Indeed, glaciation has been argued[6] to dictate large-scale deformation in the Patagonian Andes, but is probably not significant in the Bolivian Andes. The search for climatic controls on tectonics may thus require a push into glacial territory. ∎

Alison M. Anders *is in the Department of Geology, University of Illinois Urbana-Champaign, Champaign, Illinois 61820, USA.*
e-mail: amanders@illinois.edu

1. Whipple, K. X. & Gasparini, N. M. *Lithosphere* http://dx.doi.org/10.1130/L325.1 (2014).
2. Gasparini, N. M. & Whipple, K. X. *Lithosphere* http://dx.doi.org/10.1130/L322.1 (2014).
3. Wobus, C. W., Crosby, B. T. & Whipple, K. X. *J. Geophys. Res. Earth Surf.* **111,** F02017 (2006).
4. Barke, R. & Lamb, S. *Earth Planet. Sci. Lett.* **249,** 350–367 (2006).
5. Marshak, S. in *Thrust Tectonics and Hydrocarbon Systems* (ed. McClay, K. R.) 131–156 (AAPG, 2004).
6. Thomson, S. N. *et al. Nature* **467,** 313–317 (2010).
7. Nesbitt, S. W. & Anders, A. M. *Geophys. Res. Lett.* **36,** L15815 (2009).

POPULATION BIOLOGY

Fur seals signal their own decline

Data on three generations of Antarctic fur seals suggest that climate change is reducing the survival of less-fit individuals with low genetic variation, but that overall seal numbers are falling. SEE LETTER P.462

TIM COULSON & SONYA CLEGG

A little more than a century ago, humans had pushed Antarctic fur seals (*Arctocephalus gazella*) to the brink of extinction. Once hunting stopped, their numbers rose, and by the early twenty-first century their population had grown to a few million. On page 462 of this issue, Forcada and Hoffman[1] show that breeding fur seals on the beaches of South Georgia Island in the southern Atlantic are, on average, becoming more heterozygous — an indicator of individual genetic variability that is linked to improved survival and reproductive success. At face value, this seems to be good news, but a more detailed look reveals that it is a worrying symptom of a population that is again declining.

Offspring of sexually reproducing organisms receive one strand of DNA from their mother and another from their father. Some points along the two strands are the same and others are different; the more differences, the more heterozygous an individual is said to be. Higher heterozygosity frequently correlates with an individual's ability to successfully survive and reproduce — more-heterozygous individuals are fitter[2–4].

Using individual-level data on the survival, reproduction and genetics of multiple generations of fur seals coupled with data on stocks of krill (the seals' main food) and weather

conditions in the southern oceans, Forcada and Hoffman tested whether there have been changes in the size of the population and the heterozygosity of the fur seals that are associated with changing environmental conditions. They found that there has been no change in the heterozygosity of seal pups born, but that there has been an increase in the mortality rates of less-heterozygous individuals compared with those of higher heterozygosity. This means that only the more-heterozygous individuals are surviving to breeding age. But the data also show that these individuals begin breeding later in life, and breed less frequently. During the course of the study, the authors report a decline in population size of nearly 25%, and a per-generation increase in the heterozygosity of breeding females of 8.5%.

The reason that these breeding animals are not producing more-heterozygous offspring is that heterozygosity is a complex function of how different each parent is at each point along the strands of DNA they pass to their offspring. Two highly heterozygous parents may be just as likely to produce an offspring with low heterozygosity as they are to produce one that is highly heterozygous. Heterozygosity itself is not heritable and, consequently, the population cannot adapt to change through evolving levels of heterozygosity (Fig. 1). But why are we seeing elevated mortality rates in the less-heterozygotic individuals?
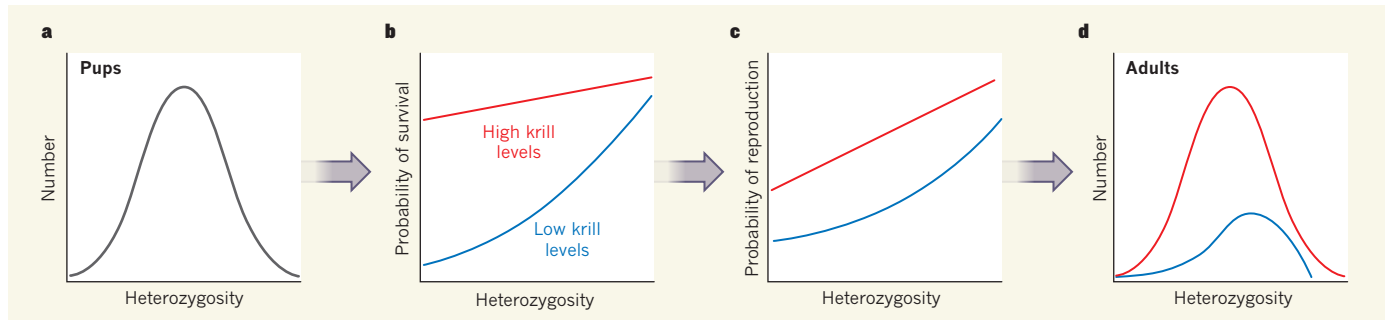
To investigate this, Forcada and Hoffman

Figure 1 | Climate change in the genes. a, The heterozygosity of fur-seal pups remains constant with time. **b, c,** By contrast, the associations of heterozygosity with survival (**b**) and reproduction rate (**c**) vary with the availability of krill (fur seals' main food), which is influenced by climate. **d,** As a consequence, in years with low krill availability (owing to high sea surface temperatures), only more-heterozygous females survive to adulthood.

constructed a simulation model of the number of breeding females, incorporating a measure of climatic variation in the southern oceans and observed variation in the availability of krill. The model provided predictions that matched well with observed fluctuations in population size, suggesting that recent declines in fur-seal numbers have been driven to a large extent by changes in weather patterns that have reduced the availability of krill. Other factors may have also affected krill availability, including changes in fishing practices[5] and increased whale numbers[6].

The authors found that a model including krill availability and climatic variation predicted observed population sizes a little better than one that included only climatic variation, which suggests that increases in fishing pressure and whale numbers may have also contributed to the recent decline in the fur-seal population. These changes have had a disproportionately greater effect on fur seals with low heterozygosity than those with higher values, leading to a simultaneous decline in population size and an increase in the heterozygosity of breeding females (Fig. 1).

Why does any of this matter? One of the biggest unknowns in predicting how our planet's weather will continue to change in the future — arguably the biggest threat that humanity faces over the coming century — is how animals, plants and the ecosystems they form will respond to changing carbon dioxide levels, and how these changes will feed back to influence the global climate. We have few studies that adequately explore the response of natural systems to environmental change, because collecting the necessary data is challenging and simple surveys of one or two aspects of a population or ecosystem are often insufficient to provide useful understanding. If Forcada and Hoffman had solely focused on the trend in heterozygosity among seals, they might have concluded that the future for the seals is bright because they are getting fitter. But the detailed, long-term, individual-based data reveal a completely different, and altogether less rosy, picture.

Fortunately, technological advances now mean that ecologists and ecosystems scientists

can gather considerably more data that are appropriate for assessing how the natural world is responding to climate change, and how these responses feed back to either decrease or accelerate rates of change. Such technology is not cheap, and investment from government and industry is required for its deployment. But once that happens, we will be able to get a much clearer picture as to whether the fur seals of South Georgia are unusual in their response to anthropogenic change, or whether such genetic and ecological dynamics are typical. Either way, it may be much harder to arrest the ongoing decline in fur seals in the twenty-first century than it was in the twentieth. ∎

Tim Coulson and **Sonya Clegg** *are in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.* **S.C.** *is also at the Griffith School of Environment and Environmental Futures Research Institute, Griffith University, Gold Coast Campus, Queensland, Australia. e-mails: timothy.coulson@zoo.ox.ac.uk; sonya.clegg@zoo.ox.ac.uk Twitter: @tncoulson*

1. Forcada, J. & Hoffman, J. I. *Nature* **511,** 462–465 (2014).
2. Amos, W. *et al. Proc. R. Soc. B* **268,** 2021–2027 (2001).
3. Chapman, J. R., Nakagawa, S., Coltman, D. W., Slate, J. & Sheldon, B. C. *Mol. Ecol.* **18,** 2746–2765 (2009).
4. Thelen, G. C. & Allendorf, F. W. *Evolution* **55,** 1180–1187 (2001).
5. Nicol, S., Foster, J. & Kawaguchi, S. *Fish Fish.* **13,** 30–40 (2012).
6. Magera, A. M., Mills Flemming, J. E., Kaschner, K., Christensen, L. B. & Lotze, H. K. *PLoS ONE* **8,** e77908 (2013).

CLIMATE SCIENCE

Cold carbon storage

Lakes that form in thawing permafrost emit substantial amounts of greenhouse gases to the atmosphere. It emerges that large quantities of carbon can also be stored in sediments at the lake bottoms. SEE LETTER P.452

SEBASTIAN SOBEK

Frozen soils in the Arctic contain more than twice the amount of carbon present as carbon dioxide in the atmosphere[1]. In the worst-case scenario, climate warming will thaw these permafrost soils and release the long-stored carbon, making it available to microbes that degrade it to the greenhouse gases CO_2 or methane[2,3]. This would further heat the atmosphere and accelerate permafrost thaw. However, it is by no means certain that the worst case is the most likely one. On page 452 of this issue, Walter Anthony *et al.*[4] show that lakes that formed in thawing permafrost thousands of years ago have accumulated vast amounts of plant remains in sediments at their bottoms, building up a large carbon sink that has had a cooling effect on climate.

Permafrost thaw is nothing new. During the most recent ice age, the majority of northeastern Siberia and Alaska was covered not by ice, but by permafrost soils called yedoma[5] up to 90 metres thick, which formed when wind-blown dust accumulated there and froze. The yedoma started to thaw some 15,000 years ago as the climate warmed, and the big chunks of ice contained within it melted to become lakes (called thaw or thermokarst lakes). As meltwater eroded the soils, large quantities of old carbon were released and degraded by aquatic microbes, leading to considerable emissions of methane gas — a process observed in present-day thaw lakes[2]. Given that rapid climate change has partially triggered carbon release in the past[6], future prospects are worrisome.

However, lakes almost always act as both carbon sources and sinks. Although CO_2

50 Years Ago

As regards technical development, very high frequency, "VHF", is undoubtedly the radio system of the future. The possibility of stereophonic broadcasting on some regular scheduled basis is just beginning to show above the horizon. Even to-day, some 6 million people who have no television depend entirely on sound for their broadcasting service, and the needs of this audience are as diverse as ever. Extensions which the British Broadcasting Corporation has in mind at present concern particularly the Light Programme and the Third Network. **From** *Nature* **25 July 1964**

100 Years Ago

The Progress of Eugenics. By Dr. C. W. Saleeby — Dr. Saleeby divides eugenics into natural or primary and nurtural or secondary. Natural eugenics is further sub-divided into positive, negative, and preventive … In treating these subjects Dr. Saleeby says, "We must be scientific or we are lost," and it is certainly true that he would have succeeded better if he had himself maintained a more scientific attitude. He falls far short of it in particular in that he appears to judge of the validity of scientific work by the conclusions it arrives at … Nevertheless, there is much contained in the book that is sensible … and this circumstance makes its faults all the more regrettable. Besides that to which allusion has already been made there are two others, first the obtrusive egotism of the writer, and secondly his habit of misrepresenting people from whom he differs in opinion. To say that "for years the chief object of the biometrical laboratory at University College has seemed to be, and now clearly is, to prove the inheritance of this or that human character is 'not Mendelian'" is little short of libellous. **From** *Nature* **23 July 1914**

Figure 1 | Thermokarst lakes in Siberia. Walter Anthony *et al.*[4] report a substantial carbon sink in sediments that formed at the bottom of lakes created from melting permafrost.

and methane are emitted to the air, sinking particles form sediment at lake bottoms and thereby bury carbon. This process is well understood, and happens in a similar way in the ocean. But most studies on the carbon balance of lakes, including thaw lakes, have focused on greenhouse-gas emission — perhaps because gas emission is more obviously linked to climate change than is mud.

Walter Anthony and colleagues used shovels to dig up to 20 metres deep into the sediments of thaw lakes in Siberia (Fig. 1). This was possible because the thaw lakes have gone through a typical series of different phases: once formed, meltwater continued to erode the frozen soil until eventually a lateral or below-ground channel opened up and drained the water from the lake basins. The sediments were suddenly exposed to cold Arctic air and froze, forming a carbon-rich permafrost soil type called alas.

The authors found that carbon accumulated at high rates in permafrost thaw lakes that had formed after the most recent ice age. On the basis of observations of present-day thaw lakes, they infer that high nutrient levels supplied from thawing permafrost boosted the growth of aquatic plants such as mosses and sedges, and that near-zero oxygen levels and low temperatures in bottom waters inhibited microbial degradation of accumulating plant remains.

By calculating gas emission and carbon burial back in time, Walter Anthony and co-workers conclude that thaw lakes acted as strong methane sources starting 15,000 years ago, but switched to become carbon sinks about 5,000 years ago. Summing up this behaviour for the past 15,000 years, it seems that thaw lakes have buried more carbon than they emitted to the atmosphere. The researchers estimate that this massive burial has resulted in

alas carbon making up about one-third of today's carbon stock in permafrost soils in the region where yedoma is the dominant soil type.

Revealing a significant carbon sink in a region that is thought to be a hazard for climate change is a spectacular achievement. But, needless to say, uncertainties in the authors' large-scale, back-in-time calculations are large. Pronounced variations in gas emissions over space and time[7,8] make generalizations challenging, even for fairly well-studied, present-day lakes; estimating gas emissions for the past 15,000 years necessarily relies on numerous assumptions.

For example, shifting environmental conditions may change the ratio of the amount of CO_2 and methane produced, as well as the extent of methane oxidation to CO_2 by aquatic microbes, thereby affecting estimates of the climate effect of gas emission. Also, the accumulated plant remains count as a carbon sink only if the plants used atmospheric CO_2 for growth, but intensive degradation of yedoma in the thaw lakes may have provided growing plants with yedoma-derived CO_2 — how much is uncertain. Irrespective of these and other uncertainties, Walter Anthony and colleagues' paper adds another layer to our picture of permafrost landscapes, and shows clearly that conclusions about ecosystem carbon balances that rely solely on atmospheric gas exchange are misleading[9,10].

But does the paper dispel the idea of massive, warming-induced carbon release from thaw lakes? Certainly not. The authors' calculations reveal that a pronounced methane-emission pulse occurred on thaw-lake formation, resulting in a climate-warming effect that lasted several thousand years. Even if sedimentary carbon burial in presently forming thaw

lakes could alleviate — or even compensate for — gas emission in the long term, waiting for the cooling effect would take more time than we can afford if we are to avoid serious effects of climate change.

The probability of a huge carbon release will depend not only on climate, but also on the degradability of alas carbon upon thaw: if we are unlucky, alas carbon degrades at similarly high rates to those reported for yedoma carbon[11], but information about this is scarce[5,12]. We need to fill fundamental gaps in our understanding of the complex dynamics of the permafrost landscape if we are to predict

the effects of climate change on such a huge carbon stock. ∎

Sebastian Sobek *is in the Department of Ecology and Genetics, Limnology, Uppsala University, 75236 Uppsala, Sweden.*
e-mail: sebastian.sobek@ebc.uu.se

1. Tarnocai, C. *et al. Glob. Biogeochem. Cycles* **23,** GB2023; http://dx.doi. org/10.1029/2008gb003327 (2009).
2. Walter, K. M., Zimov, S. A., Chanton, J. P., Verbyla, D. & Chapin, F. S. *Nature* **443,** 71–75 (2006).
3. Schuur, E. A. G. *et al. Nature* **459,** 556–559 (2009).
4. Walter Anthony, K. M. *et al.* **511,** 452–456 (2014).
5. Zimov, S. A. *et al. Geophys. Res. Lett.* **33,** L20502; http://dx.doi.org/10.1029/2006gl027484 (2006).
6. Walter, K. M., Edwards, M. E., Grosse, G., Zimov, S. A. & Chapin, F. S. III *Science* **318,** 633–636 (2007).
7. Rantakari, M. & Kortelainen, P. *Glob. Change Biol.* **11,** 1368–1380 (2005).
8. Bastviken, D., Cole, J., Pace, M. & Tranvik, L. *Glob. Biogeochem. Cycles* **18,** GB4009; http://dx.doi.org/10.1029/2004GB002238 (2004).
9. Ciais, P. *et al. Biogeosciences* **5,** 1259–1271 (2008).
10. Mendonça, R. *et al. Nature Geosci.* **5,** 838–840 (2012).
11. Vonk, J. E. *et al. Geophys. Res. Lett.* **40,** 2689–2693 (2013).
12. Knoblauch, C., Beer, C., Sosnin, A., Wagner, D. & Pfeiffer, E.-M. *Glob. Change Biol.* **19,** 1160–1172 (2013).

GENE REGULATION

Fine-tuned amplification in cells

The transcription factor Myc has been posited to cause a cell-wide increase in gene expression. But two studies show that Myc, when modulated by other transcription factors, can amplify select targets. SEE LETTERS P.483 AND P.488

CHI V. DANG

The Myc protein can both stimulate and repress RNA synthesis; however, the means by which it exerts these opposing effects are unclear. This gap in our knowledge is of interest to many researchers, because *Myc* gene expression is itself strictly regulated during normal cellular proliferation, but often deregulated in human cancers. Moreover, deregulated *Myc* expression can initiate cancer in mice[1], suggesting that the way in which Myc regulates gene expression in healthy cells must be different from how it does so in cancer. Two papers[2,3] in this issue provide insights into Myc's differing roles in normal and cancerous cells and show that, contrary to recent hypotheses, Myc can selectively regulate gene expression.

Transcription factors such as Myc typically regulate gene expression by binding to target regulatory DNA regions to inhibit or promote transcription. For many years, Myc was thought to regulate specific genes involved in metabolism and cell growth[1]. However, in 2012 it emerged[4] that Myc could stimulate production of almost all messenger RNA transcripts in a cell when cell size was taken into account (which is relevant because Myc can increase cell size and, generally speaking, there is more transcription in larger cells). The same year, two papers[5,6] questioned the existence of specific Myc target genes, postulating instead that Myc directly upregulates all transcripts from genes poised for transcription in a given cell, seemingly without any specificity — a theory known as the amplifier model (Fig. 1a).
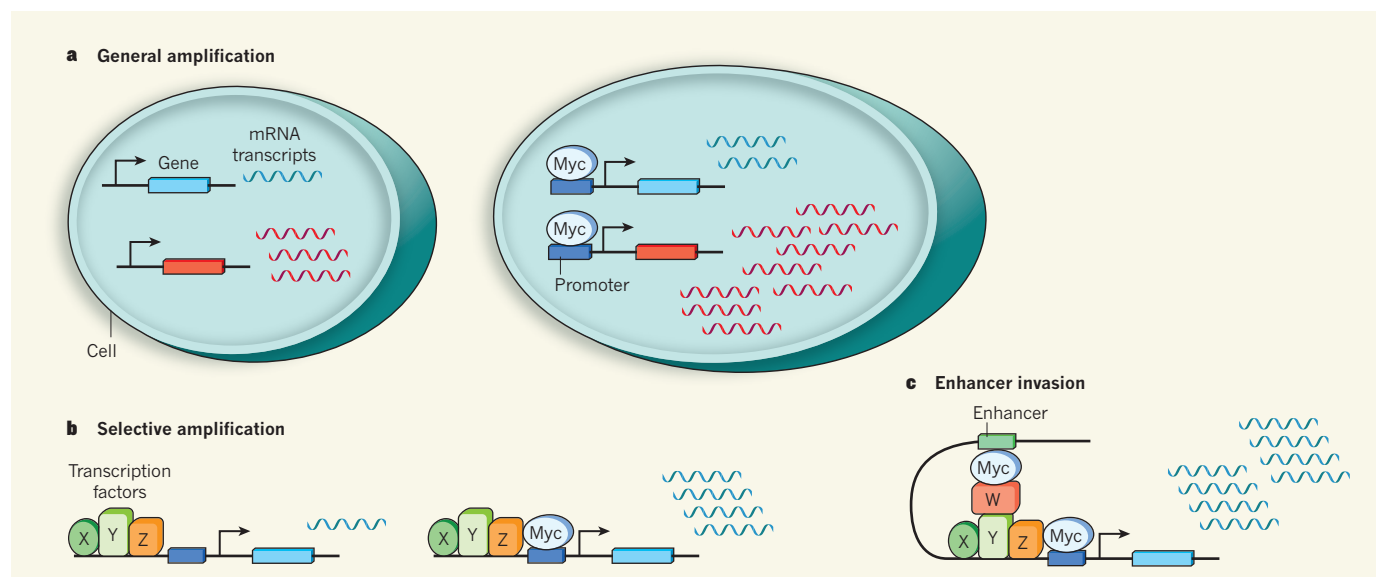

Figure 1 | Models of Myc function. The transcription factor Myc binds to regulatory promoter regions of DNA to control gene expression, but there are different models for how Myc exerts its controls. **a,** In some cell types, Myc expression increases cell size, leading to general amplification of the expression of almost all genes, which produces more messenger RNA (mRNA) transcripts. **b,** Sabò *et al.*[2] and Walz *et al.*[3] report that Myc can selectively amplify specific genes (or repress them, not shown) by interacting with other transcription factors. **c,** Sabò and colleagues show that abnormally high levels of Myc upregulate gene expression by invading enhancer sites that regulate distant genes. Myc-bound enhancers promote gene expression through chromosome looping, which brings distant sections of DNA close together.

Sabò *et al.*[2] (page 488) and Walz *et al.*[3] (page 483) now offer an alternative to the amplifier model, returning to the idea that Myc can regulate specific genes. Both groups studied cell lines in which *Myc* expression did not discernibly alter cell size, meaning that changes in gene transcription observed when *Myc* was induced might highlight genes directly targeted by Myc. The groups combined these assays with a mapping technique that pinpoints all the sites in the genome at which Myc binds to regulatory promoter regions adjacent to genes — an indication that Myc is directly regulating transcription of the nearby gene.

Walz and colleagues used human cell lines in which *Myc* expression could be experimentally manipulated, and found that several previously documented[1] putative target genes were amplified by Myc. Sabò and co-workers used a similar approach in a mouse cell line and observed that Myc activated specific sets of genes, including those involved in metabolism and assembly of the ribosome (the molecular machinery required for protein synthesis).

Interestingly, Sabò *et al.* also found that, as proposed by the amplifier model, *Myc* expression could upregulate mRNA production globally. However, this seemed to be an indirect effect, which might arise owing to the nature of the targets that Myc regulates — proteins involved in nucleotide synthesis, for example, can exert cell-wide effects on RNA and DNA synthesis. In agreement with this theory, a paper published earlier this year[7] provided molecular evidence of a translational pathway involved in nucleotide biosynthesis that can be directly activated by Myc, leading to global mRNA amplification.

The amplifier-model papers[5,6] question whether Myc can directly suppress transcription, because induction of *Myc* does not cause a relative decrease in the level of transcription from individual genes. However, both current papers define genes directly repressed by Myc. Specifically, Walz and co-workers show that Myc directly binds the transcription factor Myc-interacting zinc-finger 1 (MIZ1), and globally suppresses MIZ1 target genes. This observation underscores the role of other transcription factors in modulating the action of Myc on target genes in different cell and tissue contexts (Fig. 1b). Indeed, genes with multiple transcription factors bound to them are more likely to be directly responsive to Myc in embryonic stem cells[8]. In addition, in-depth bioinformatic analyses[9] have identified transcription-factor-binding motifs that are enriched in the regulatory regions of some Myc target genes.

One interesting aspect of Myc biology is that normal expression can drive embryonic development and tissue repair, but deregulated *Myc* promotes cancerous transformation and reprogramming of adult cells to a 'reset' stem-cell state[1]. The amplifier model does not attempt to explicitly explain cellular changes driven by abnormal *Myc* expression, but it is feasible that the model might reflect what happens in normal cells, in which Myc could directly bind to and amplify thousands of genes to regulate growth and proliferation. However, the degree of induction of specific genes by Myc is dictated by more than whether each is poised for transcription — chromosome structure, Myc–DNA binding affinity and other transcription factors that co-regulate those genes all have a part to play. Amplified expression of specific Myc target genes must result in appropriate relative levels of different mRNAs so that proteins are produced in precise amounts for normal cellular function, as has recently been documented[10] in the bacterium *Escherichia coli*.

Two points on which the current work and the amplifier model agree are that, when *Myc* is abnormally expressed, there is a correlative change in the degree of Myc–promoter binding; and that, when Myc is abnormally abundant, it can also bind to active enhancers[2,5] — regulatory DNA sequences that change gene expression from a long distance (Fig. 1c). The ability of deregulated Myc to contribute to the reprogramming of adult cells to stem cells seems to be attributable to enhancer invasion, which allows stem-cell factors to bind enhancers early in the process of reprogramming[11].

The role of high levels of Myc in enhancer regulation points to possible mechanisms for the biological differences in cells with normal or deregulated *Myc* expression. The nonlinear amplification of gene expression through enhancer invasion by activated Myc in cancer could cause deregulated expression of mRNAs, which would result in metabolic and protein imbalances on a cell-wide level. Such alterations cause oxidative stress and cellular-stress responses[12], thereby contributing to cell death or tumour development and progression. ∎

Chi V. Dang *is at the Abramson Cancer Center, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.*
e-mail: dangvchi@exchange.upenn.edu

1. Dang, C. V. *Cell* **149,** 22–35 (2012).
2. Sabò, A. *et al. Nature* **511,** 488–492 (2014).
3. Walz, S. *et al. Nature* **511,** 483–487 (2014).
4. Lovén, J. *et al. Cell* **151,** 476–482 (2012).
5. Lin, C. Y. *et al. Cell* **151,** 56–67 (2012).
6. Nie, Z. *et al. Cell* **151,** 68–79 (2012).
7. Cunningham, J. T., Moreno, M. V., Lodi, A., Ronen, S. M. & Ruggero, D. *Cell* **157,** 1088–1103 (2014).
8. Kim, J. *et al. Cell* **143,** 313–324 (2010).
9. Elkon, R. *et al. Nucleic Acids Res.* **32,** 4955–4961 (2004).
10. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. *Cell* **157,** 624–635 (2014).
11. Soufi, A., Donahue, G. & Zaret, K. S. *Cell* **151,** 994–1004 (2012).
12. Hart, L. S. *et al. J. Clin. Invest.* **122,** 4621–4634 (2012).

This article was published online on 9 July 2014.

Siphoning spins

Topological insulators are materials known for their remarkable electron–transport properties. They now emerge as excellent sources of electron spins for manipulating tiny magnets. SEE LETTER P.449

JOO-VON KIM

Conventional electronic devices use the charge of moving electrons in a circuit to perform specific tasks. For nearly 30 years, scientists have been striving to exploit another property of the electron — its spin — in electronic circuitry. Electronic currents based on the spin of electrons offer the potential for new functionalities, but generating and using such spin currents has remained a challenge. On page 449 of this issue, Mellnik *et al.*[1] report that the spin currents of an exotic class of materials called topological insulators can produce huge effects in thin magnetic films. The authors find that these currents exert larger torques on magnetic elements than do those produced in metals, paving the way for magnetic devices with improved energy efficiency for information storage and processing.

Much of the effort behind spin-based electronics, or 'spintronics', is motivated by the challenges involved in reducing the scale of microelectronic components, such as logic gates. Two problems associated with this are the detrimental heating that accompanies conventional electric currents, and the energy consumption involved in maintaining components in their idle state. The first problem can be tackled by taking advantage of the flow of spins, rather than charges, because pure spin currents do not generate heat. The second might be addressed by incorporating 'non-volatile' elements, such as magnetic memories, into logic devices, so that the devices can be powered down while retaining their logic state. Combining spin currents with magnetic materials therefore offers a promising route to overcome these problems.

Ferromagnetic metals are at the heart of spintronics, because electron flow in these materials is naturally spin-polarized — that is, one spin population (for example, 'spin-up'

electrons) is greater than the other ('spin-down' electrons). The spin polarization allows information about the ferromagnetic state to be transmitted by the electron current. By contrast, currents in normal metals are naturally unpolarized. But by combining ferromagnetic and normal metals in multilayered thin films, it is possible to produce spin-dependent effects such as giant magnetoresistance. In this phenomenon, the electrical resistance of the film is determined by how the aggregate magnetic moments that are associated with the electron spins in each ferromagnetic layer are aligned with respect to each other.

In their study, Mellnik and colleagues explore a related phenomenon in which the magnetic moments of a ferromagnet are excited by an effect called spin-transfer torque[2,3]. To understand this effect, consider the more familiar example of excitation by a magnetic field. If we imagine the moments in a ferromagnet as being represented by an ensemble of tiny compass needles, then the application of a magnetic field will tend to align the needles along the direction of the field by exerting a torque on them. Spin-transfer torques work in a similar way, but arise instead from the interaction between the magnetic moments and a spin-polarized current. In usual multilayer geometries, these currents are generated either in a second ferromagnetic layer, which acts to polarize the current, or in a heavy-metal underlayer through the spin Hall effect — the separation of electrons of opposite spin orientation in an electric field.

Mellnik and co-workers' key finding is that spin currents at the surface of a topological insulator can produce much greater spin-transfer effects than those produced at similar current densities by a metal. Topological insulators are good conductors on their surfaces but remain insulators in the bulk[4–6], much like a copper-plated slab of glass. They have generated great excitement among condensed-matter physicists over the past few years because of their electronic properties. The interesting feature here is that the currents in a topological insulator are strongly spin-polarized, leading to the accumulation of spins of a specific orientation at its surface as the charge flows along it. This means that spins can be 'siphoned' off the surface of the insulator if a metal comes into contact with it.
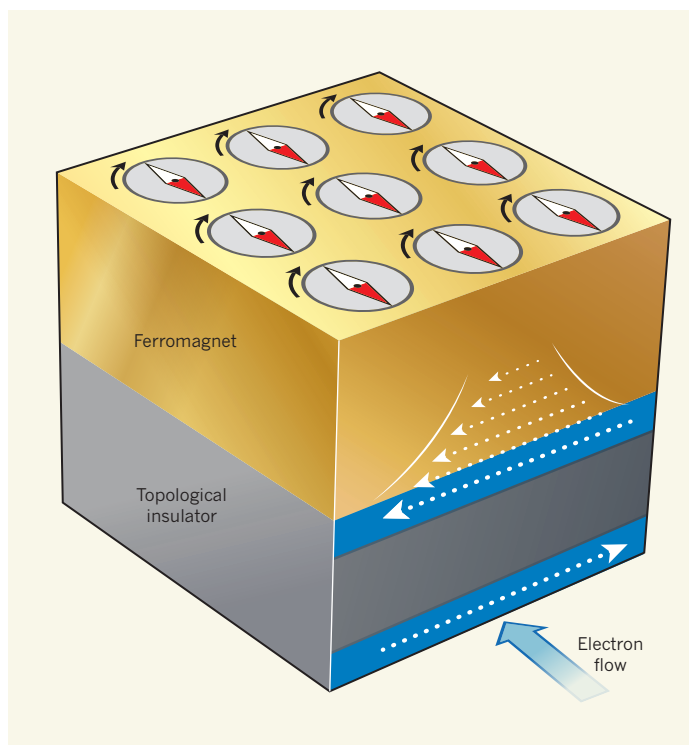

Figure 1 | Taking the spin off a topological insulator's surface. Mellnik *et al.*[1] have applied an electric current to a bilayer comprising a nickel–iron ferromagnet and a bismuth selenide topological insulator, and observed how the diffusion of electron spins from the insulator into the ferromagnet affects the magnetic moments of the ferromagnet. The encircled, red-tipped arrows ('compass needles') represent the magnetic moments of the ferromagnetic layer. The blue areas are cross-sections of the regions of the topological insulator through which electron charges can flow with little resistance — current flows only along the surface of the topological insulator. The dashed arrows represent accumulated spins of a specific orientation arising from the electron flow. The solid white lines represent the spin diffusion into the ferromagnet, which is described by an exponential decay of the spin accumulation from the interface. The spin diffusion exerts torques on the magnetic moments, causing them to rotate from their equilibrium orientation.

This is precisely what Mellnik *et al.* demonstrate. Their experiment involves a thin nickel–iron ferromagnet in contact with a film of the topological insulator bismuth selenide (Fig. 1). By applying an alternating current to this bilayer, the spins diffusing from the topological insulator into the ferromagnet drive the magnetic moments of the ferromagnet into resonance — that is, their precession is greatly amplified at a specific frequency of the alternating current. The amplitude of this resonance provides information about the strength of the spin-transfer torques. Control experiments and comparison with previous work involving heavy metals[7], in which spin currents are produced in a similar geometry, show that topological insulators can provide up to a 40-fold increase in the spin torque per unit electron-current density. This is quite remarkable, given that the authors found that spin conductivities of the surface currents in the topological insulator were comparable to those for currents in heavy metals, in which spin currents flowing perpendicular to the films are generated by the spin Hall effect.

Although the large spin torques observed are unprecedented, a few unresolved issues remain. For example, it is difficult to determine experimentally how much of the electron current actually flows in the topological insulator or how insulating the bulk of this material really is. These issues lead to the authors' reported uncertainty in the spin-torque angle — a quantity that characterizes the spin-transfer torque and is useful for applications. Further elucidation of the proposed spin-torque mechanism, which involves accumulated spins diffusing into the ferromagnet, could be verified in future experiments by studying a large range of ferromagnet thicknesses. This would also shed light on whether spin currents due to the spin Hall effect are present.

Regardless of the origin of the observed spin torques, Mellnik and colleagues' study opens up the prospect of incorporating topological insulators into real spintronic devices. This is possibly the most important outcome, because the study shows that the basic features of topological insulators are sufficiently robust even in imperfect films, and that pristine samples are not necessary to achieve spectacular effects. Recent work[8] on topological-insulator heterostructures doped with magnetic impurities also represents a step in this direction. If the history of giant magnetoresistance can serve as a guide — barely a decade passed from its discovery in 'model' (epitaxially grown) systems to its widespread use in applications with imperfect sputtered films — then we may not have long to wait before topological insulators feature in spintronic devices. ■

Joo-Von Kim *is at the Institut d'Electronique Fondamentale, CNRS, and Université Paris-Sud, 91405 Orsay, France.*
e-mail: joo-von.kim@u-psud.fr

1. Mellnik, A. R. *et al. Nature* **511,** 449–451 (2014).
2. Berger, L. *Phys. Rev. B* **54,** 9353–9358 (1996).
3. Slonczewski, J. C. *J. Magn. Magn. Mater.* **159,** L1–L7 (1996).
4. Moore, J. E. *Nature* **464,** 194–198 (2010).
5. Hasan, M. Z. & Kane, C. L. *Rev. Mod. Phys.* **82,** 3045–3067 (2010).
6. Qi, X.-L. & Zhang, S.-C. *Rev. Mod. Phys.* **83,** 1057–1110 (2011).
7. Liu, L. *et al. Science* **336,** 555–558 (2012).
8. Fan, Y. *et al. Nature Mater.* **13,** 699–704 (2014).

Biological insights from 108 schizophrenia–associated genetic loci

Schizophrenia Working Group of the Psychiatric Genomics Consortium*

Schizophrenia is a highly heritable disorder. Genetic risk is conferred by a large number of alleles, including common alleles of small effect that might be detected by genome-wide association studies. Here we report a multi-stage schizophrenia genome-wide association study of up to 36,989 cases and 113,075 controls. We identify 128 independent associations spanning 108 conservatively defined loci that meet genome-wide significance, 83 of which have not been previously reported. Associations were enriched among genes expressed in brain, providing biological plausibility for the findings. Many findings have the potential to provide entirely new insights into aetiology, but associations at *DRD2* and several genes involved in glutamatergic neurotransmission highlight molecules of known and potential therapeutic relevance to schizophrenia, and are consistent with leading pathophysiological hypotheses. Independent of genes expressed in brain, associations were enriched among genes expressed in tissues that have important roles in immunity, providing support for the speculated link between the immune system and schizophrenia.

Schizophrenia has a lifetime risk of around 1%, and is associated with substantial morbidity and mortality as well as personal and societal costs[1–3]. Although pharmacological treatments are available for schizophrenia, their efficacy is poor for many patients[4]. All available antipsychotic drugs are thought to exert their main therapeutic effects through blockade of the type 2 dopaminergic receptor[5,6] but, since the discovery of this mechanism over 60 years ago, no new antipsychotic drug of proven efficacy has been developed based on other target molecules. Therapeutic stasis is in large part a consequence of the fact that the pathophysiology of schizophrenia is unknown. Identifying the causes of schizophrenia is therefore a critical step towards improving treatments and outcomes for those with the disorder.

High heritability points to a major role for inherited genetic variants in the aetiology of schizophrenia[7,8]. Although risk variants range in frequency from common to extremely rare[9], estimates[10,11] suggest half to a third of the genetic risk of schizophrenia is indexed by common alleles genotyped by current genome-wide association study (GWAS) arrays. Thus, GWAS is potentially an important tool for understanding the biological underpinnings of schizophrenia.

To date, around 30 schizophrenia-associated loci[10–23] have been identified through GWAS. Postulating that sample size is one of the most important limiting factors in applying GWAS to schizophrenia, we created the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC). Our primary aim was to combine all available schizophrenia samples with published or unpublished GWAS genotypes into a single, systematic analysis[24]. Here we report the results of that analysis, including at least 108 independent genomic loci that exceed genome-wide significance. Some of the findings support leading pathophysiological hypotheses of schizophrenia or targets of therapeutic relevance, but most of the findings provide new insights.

108 independent associated loci

We obtained genome-wide genotype data from which we constructed 49 ancestry matched, non-overlapping case-control samples (46 of European and three of east Asian ancestry, 34,241 cases and 45,604 controls) and 3 family-based samples of European ancestry (1,235 parent affected-offspring trios) (Supplementary Table 1 and Supplementary Methods).

These comprise the primary PGC GWAS data set. We processed the genotypes from all studies using unified quality control procedures followed by imputation of SNPs and insertion-deletions using the 1000 Genomes Project reference panel[25]. In each sample, association testing was conducted using imputed marker dosages and principal components (PCs) to control for population stratification. The results were combined using an inverse-variance weighted fixed effects model[26]. After quality control (imputation INFO score $\geq$ 0.6, MAF $\geq$ 0.01, and successfully imputed in $\geq$ 20 samples), we considered around 9.5 million variants. The results are summarized in Fig. 1. To enable acquisition of large samples, some groups ascertained cases via clinician diagnosis rather than a research-based assessment and provided evidence of the validity of this approach (Supplementary Information)[11,13]. Post hoc analyses revealed the pattern of effect sizes for associated loci was similar across different assessment methods and modes of ascertainment (Extended Data Fig. 1), supporting our *a priori* decision to include samples of this nature.

For the subset of linkage-disequilibrium-independent single nucleotide polymorphisms (SNPs) with $P < 1 \times 10^{-6}$ in the meta-analysis, we next obtained results from deCODE genetics (1,513 cases and 66,236 controls of European ancestry). We define linkage-disequilibrium-independent SNPs as those with low linkage disequilibrium ($r^2 < 0.1$) to a more significantly associated SNP within a 500-kb window. Given high linkage disequilibrium in the extended major histocompatibility complex (MHC) region spans ∼8 Mb, we conservatively include only a single MHC SNP to represent this locus. The deCODE data were then combined with those from the primary GWAS to give a data set of 36,989 cases and 113,075 controls. In this final analysis, 128 linkage-disequilibrium-independent SNPs exceeded genome-wide significance ($P \leq 5 \times 10^{-8}$) (Supplementary Table 2).

As in meta-analyses of other complex traits which identified large numbers of common risk variants[27,28], the test statistic distribution from our GWAS deviates from the null (Extended Data Fig. 2). This is consistent with the previously documented polygenic contribution to schizophrenia[10,11]. The deviation in the test statistics from the null ($\lambda_{GC} = 1.47$, $\lambda_{1000} = 1.01$) is only slightly less than expected ($\lambda_{GC} = 1.56$) under a polygenic model given fully informative genotypes, the current sample size, and the lifetime risk and heritability of schizophrenia[29].

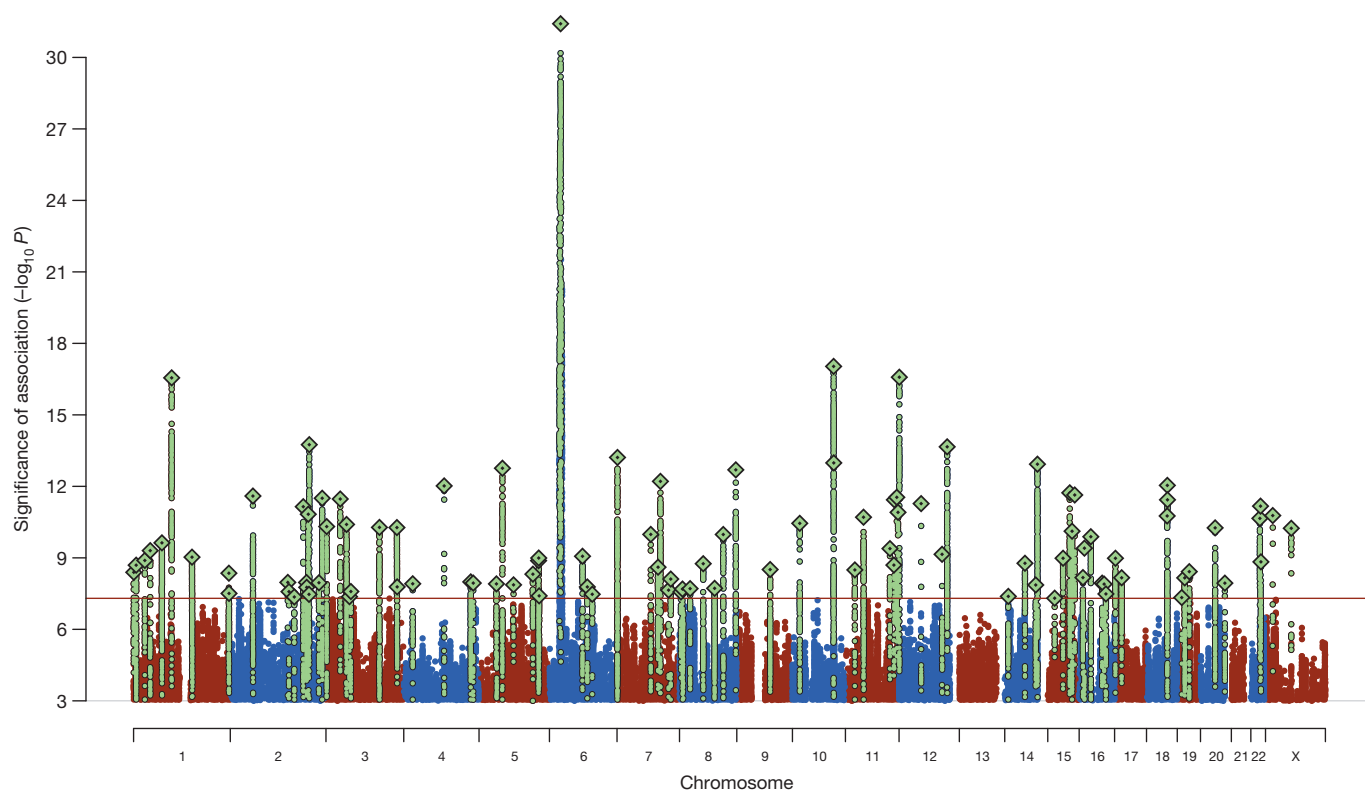*A list of authors and affiliations appears at the end of the paper.

Figure 1 | Manhattan plot showing schizophrenia associations. Manhattan plot of the discovery genome-wide association meta-analysis of 49 case control samples (34,241 cases and 45,604 controls) and 3 family based association studies (1,235 parent affected-offspring trios). The *x* axis is chromosomal position and the *y* axis is the significance ($-\log_{10} P$; 2-tailed) of association derived by logistic regression. The red line shows the genome-wide significance level (5×10^{-8}). SNPs in green are in linkage disequilibrium with the index SNPs (diamonds) which represent independent genome-wide significant associations.

Additional lines of evidence allow us to conclude the deviation between the observed and null distributions in our primary GWAS indicates a true polygenic contribution to schizophrenia. First, applying a novel method[30] that uses linkage disequilibrium information to distinguish between the major potential sources of test statistic inflation, we found our results are consistent with polygenic architecture but not population stratification (Extended Data Fig. 3). Second, the schizophrenia-associated alleles at 78% of 234 linkage-disequilibrium-independent SNPs exceeding $P < 1 \times 10^{-6}$ in the case-control GWAS were again overrepresented in cases in the independent samples from deCODE. This degree of consistency between the case-control GWAS and the replication data is highly unlikely to occur by chance ($P = 6 \times 10^{-19}$). The tested alleles surpassed the $P < 10^{-6}$ threshold in our GWAS before we added either the trios or deCODE data to the meta-analysis. This trend test is therefore independent of the primary case-control GWAS. Third, analysing the 1,235 parent-proband trios, we again found excess transmission of the schizophrenia-associated allele at 69% of the 263 linkage-disequilibrium-independent SNPs with $P < 1 \times 10^{-6}$ in the case-control GWAS. This is again unlikely to occur by chance ($P = 1 \times 10^{-9}$) and additionally excludes population stratification as fully explaining the associations reaching our threshold for seeking replication. Fourth, we used the trios trend data to estimate the expected proportion of true associations at $P < 1 \times 10^{-6}$ in the discovery GWAS, allowing for the fact that half of the index SNPs are expected to show the same allelic trend in the trios by chance, and that some true associations will show opposite trends given the limited number of trio samples (Supplementary Methods). Given the observed trend test results, around 67% (95% confidence interval: 64–73%) or $n = 176$ of the associations in the scan at $P < 1 \times 10^{-6}$ are expected to be true, and therefore the number of associations that will ultimately be validated from this set of SNPs will be considerably more than those that now meet genome-wide significance. Taken together, these analyses indicate that the observed deviation

of test statistics from the null primarily represents polygenic association signal and the considerable excess of associations at the tail of extreme significance largely correspond to true associations.

Independently associated SNPs do not translate to well-bounded chromosomal regions. Nevertheless, it is useful to define physical boundaries for the SNP associations to identify candidate risk genes. We defined an associated locus as the physical region containing all SNPs correlated at $r^2 > 0.6$ with each of the 128 index SNPs. Associated loci within 250 kb of each other were merged. This resulted in 108 physically distinct associated loci, 83 of which have not been previously implicated in schizophrenia and therefore harbour potential new biological insights into disease aetiology (Supplementary Table 3; regional plots in Supplementary Fig. 1). The significant regions include all but 5 loci previously reported to be genome-wide significant in large samples (Supplementary Table 3).

Characterization of associated loci

Of the 108 loci, 75% include protein-coding genes (40%, a single gene) and a further 8% are within 20 kb of a gene (Supplementary Table 3). Notable associations relevant to major hypotheses of the aetiology and treatment of schizophrenia include *DRD2* (the target of all effective antipsychotic drugs) and many genes (for example, *GRM3, GRIN2A, SRR, GRIA1*) involved in glutamatergic neurotransmission and synaptic plasticity. In addition, associations at *CACNA1C, CACNB2* and *CACNA1I*, which encode voltage-gated calcium channel subunits, extend previous findings implicating members of this family of proteins in schizophrenia and other psychiatric disorders[11,13,31,32]. Genes encoding calcium channels, and proteins involved in glutamatergic neurotransmission and synaptic plasticity have been independently implicated in schizophrenia by studies of rare genetic variation[33–35], suggesting convergence at a broad functional level between studies of common and rare genetic variation. We highlight in the Supplementary Discussion genes of particular interest within associated loci with respect to current hypotheses of schizophrenia

aetiology or treatment (although we do not imply that these genes are necessarily the causal elements).

For each of the schizophrenia-associated loci, we identified a credible causal set of SNPs (for definition, see Supplementary Methods)[36]. In only 10 instances (Supplementary Table 4) was the association signal credibly attributable to a known non-synonymous exonic polymorphism. The apparently limited role of protein-coding variants is consistent both with exome sequencing findings[33] and with the hypothesis that most associated variants detected by GWAS exert their effects through altering gene expression rather than protein structure[37,38] and with the observation that schizophrenia risk loci are enriched for expression quantitative trait loci (eQTL)[39].

To try to identify eQTLs that could explain associations with schizophrenia, we merged the credible causal set of SNPs defined above with eQTLs from a meta-analysis of human brain cortex eQTL studies ($n = 550$) and an eQTL study of peripheral venous blood ($n = 3,754$)[40] (Supplementary Methods). Multiple schizophrenia loci contained at least one eQTL for a gene within 1 Mb of the locus (Supplementary Table 4). However, in only 12 instances was the eQTL plausibly causal (two in brain, and nine in peripheral blood, one in both). This low proportion suggests that if most risk variants are regulatory, available eQTL catalogues do not yet provide power, cellular specificity, or developmental diversity to provide clear mechanistic hypotheses for follow-up experiments.

The brain and immunity

To further explore the regulatory nature of the schizophrenia associations, we mapped the credible sets ($n = 108$) of causal variants onto sequences with epigenetic markers characteristic of active enhancers in 56 different tissues and cell lines (Supplementary Methods). Schizophrenia associations were significantly enriched at enhancers active in brain (Fig. 2) but not in tissues unlikely to be relevant to schizophrenia (for example, bone, cartilage, kidney and fibroblasts). Brain tissues used to define enhancers consist of heterogeneous populations of cells. Seeking greater specificity, we contrasted genes enriched for expression in neurons and glia using mouse ribotagged lines[41]. Genes with strong expression in multiple cortical and striatal neuronal lineages were enriched for associations, providing support for an important neuronal pathology in schizophrenia (Extended Data Fig. 4) but this is not statistically more significant than, or exclusionary of, contributions from other lineages[42].

Schizophrenia associations were also strongly enriched at enhancers that are active in tissues with important immune functions, particularly B-lymphocyte lineages involved in acquired immunity (CD19 and CD20 lines, Fig. 2). These enrichments remain significant even after excluding the extended MHC region and regions containing brain enhancers (enrichment P for CD20 $< 10^{-6}$), demonstrating that this finding is not an artefact of correlation between enhancer elements in different tissues and not driven by the strong and diffuse association at the extended MHC. Epidemiological studies have long hinted at a role for immune dysregulation in schizophrenia, the present findings provide genetic support for this hypothesis[43].

To develop additional biological hypotheses beyond those that emerge from inspection of the individual loci, we further undertook a limited mining of the data through gene-set analysis. However, as there is no consensus methodology by which such analyses should be conducted, nor an established optimal significance threshold for including loci, we sought to be conservative, using only two of the many available approaches[44,45] and restricting analyses to genes within genome-wide significant loci. Neither approach identified gene-sets that were significantly enriched for associations after correction for the number of pathways tested (Supplementary Table 5) although nominally significantly enrichments were observed among several predefined candidate pathways (Extended Data Table 1). A fuller exploratory analysis of the data will be presented elsewhere.

Overlap with rare mutations

CNVs associated with schizophrenia overlap with those associated with autism spectrum disorder (ASD) and intellectual disability[9], as do genes
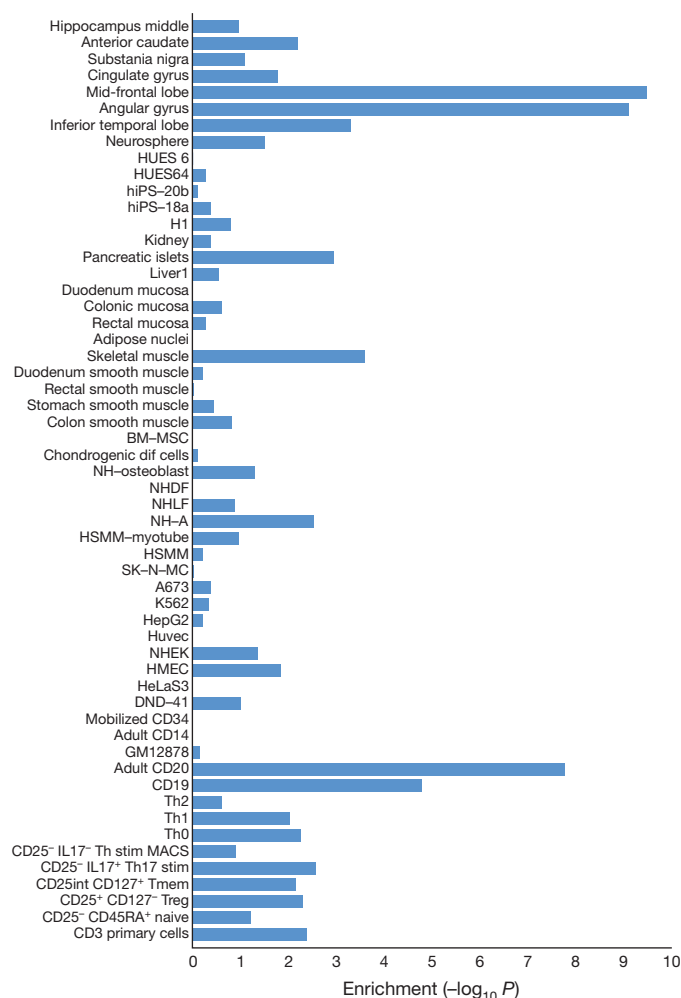

Figure 2 | Enrichment in enhancers of credible SNPs. Cell and tissue type specific enhancers were identified using ChIP-seq data sets (H3K27ac signal) from 56 cell line and tissue samples (y axis). We defined cell and tissue type enhancers as the top 10% of enhancers with the highest ratio of reads in that cell or tissue type divided by the total number of reads. Enrichment of credible causal associated SNPs from the schizophrenia GWAS was compared with frequency matched sets of 1000 Genomes SNPs (Supplementary Methods). The x axis is the $-\log_{10} P$ for enrichment. P values are uncorrected for the number of tissues or cells tested. A $-\log_{10} P$ of roughly 3 can be considered significant after Bonferroni correction. Descriptions of cell and tissue types at the Roadmap Epigenome website (http://www.roadmapepigenomics.org).

with deleterious *de novo* mutations[34]. Here we find significant overlap between genes in the schizophrenia GWAS associated intervals and those with *de novo* non-synonymous mutations in schizophrenia ($P = 0.0061$) (Extended Data Table 2), suggesting that mechanistic studies of rare genetic variation in schizophrenia will be informative for schizophrenia more widely. We also find evidence for overlap between genes in schizophrenia GWAS regions and those with *de novo* non-synonymous mutations in intellectual disability ($P = 0.00024$) and ASD ($P = 0.035$), providing further support for the hypothesis that these disorders have partly overlapping pathophysiologies[9,34].

Polygenic risk score profiling

Previous studies have shown that risk profile scores (RPS) constructed from alleles showing modest association with schizophrenia in a discovery GWAS can predict case-control status in independent samples, albeit with low sensitivity and specificity[10,11,16]. This finding was robustly confirmed in the present study. The estimate of Nagelkerke R^2 (a measure of variance in case-control status explained) depends on the specific target data set and threshold (P_T) for selecting risk alleles for RPS

analysis (Extended Data Fig. 5 and 6a). However, using the same target sample as earlier studies and $P_T = 0.05$, R^2 is now increased from 0.03 (ref. 10) to 0.184 (Extended Data Fig. 5). Assuming a liability-threshold model, a lifetime risk of 1%, independent SNP effects, and adjusting for case-control ascertainment, RPS now explains about 7% of variation on the liability scale[46] to schizophrenia across the samples (Extended Data Fig. 6b), about half of which (3.4%) is explained by genome-wide significant loci.

We also evaluated the capacity of RPS to predict case-control status using a standard epidemiological approach to a continuous risk factor. We illustrate this in three samples, each with different ascertainment schemes (Fig. 3). The Danish sample is population-based (that is, inpatient and outpatient facilities), the Swedish sample is based on all cases hospitalized for schizophrenia in Sweden, and the Molecular Genetics of Schizophrenia (MGS) sample was ascertained specially for genetic studies from clinical sources in the US and Australia. We grouped individuals into RPS deciles and estimated the odds ratios for affected status for each decile with reference to the lowest risk decile. The odds ratios increased with greater number of schizophrenia risk alleles in each sample, maximizing for the tenth decile in all samples: Denmark 7.8 (95% confidence interval (CI): 4.4–13.9), Sweden 15.0 (95% CI: 12.1–18.7) and MGS 20.3 (95% CI: 14.7–28.2). Given the need for measures that index liability to schizophrenia[47,48], the ability to stratify individuals by RPS offers new opportunities for clinical and epidemiological research. Nevertheless, we stress that the sensitivity and specificity of RPS do not support its use as a predictive test. For example, in the Danish epidemiological sample, the area under the receiver operating curve is only 0.62 (Extended Data Fig. 6c, Supplementary Table 6).

Finally, seeking evidence for non-additive effects on risk, we tested for statistical interaction between all pairs of 125 autosomal SNPs that reached genome-wide significance. P values for the interaction terms were distributed according to the null, and no interaction was significant after correction for multiple comparisons. Thus, we find no evidence for epistatic or non-additive effects between the significant loci (Extended Data Fig. 7). It is possible that such effects could be present between other loci, or occur in the form of higher-order interactions.

Discussion

In the largest (to our knowledge) molecular genetic study of schizophrenia, or indeed of any neuropsychiatric disorder, ever conducted, we demonstrate the power of GWAS to identify large numbers of risk loci. We show that the use of alternative ascertainment and diagnostic schemes designed to rapidly increase sample size does not inevitably introduce a crippling degree of heterogeneity. That this is true for a phenotype like schizophrenia, in which there are no biomarkers or supportive diagnostic tests, provides grounds to be optimistic that this approach can be successfully applied to GWAS of other clinically defined disorders.

We further show that the associations are not randomly distributed across genes of all classes and function; rather they converge upon genes that are expressed in certain tissues and cellular types. The findings include molecules that are the current, or the most promising, targets for therapeutics, and point to systems that align with the predominant aetiological hypotheses of the disorder. This suggests that the many novel findings we report also provide an aetiologically relevant foundation for mechanistic and treatment development studies. We also find overlap between genes affected by rare variants in schizophrenia and those within GWAS loci, and broad convergence in the functions of some of the clusters of genes implicated by both sets of genetic variants, particularly genes related to abnormal glutamatergic synaptic and calcium channel function. How variation in these genes impact function to increase risk for schizophrenia cannot be answered by genetics, but the overlap strongly suggests that common and rare variant studies are complementary rather than antagonistic, and that mechanistic studies driven by rare genetic variation will be informative for schizophrenia.

Figure 3 | Odds ratio by risk score profile. Odds ratio for schizophrenia by risk score profile (RPS) decile in the Sweden (Sw1-6), Denmark (Aarhus), and Molecular Genetics of Schizophrenia studies (Supplementary Methods). Risk alleles and weights were derived from 'leave one out' analyses in which those samples were excluded from the GWAS meta-analysis (Supplementary Methods). The threshold for selecting risk alleles was $P_T < 0.05$. The RPS were converted to deciles (1 = lowest, 10 = highest RPS), and nine dummy variables created to contrast deciles 2-10 to decile 1 as the reference. Odds ratios and 95% confidence intervals (bars) were estimated using logistic regression with PCs to control for population stratification.

1. Saha, S., Chant, D. & McGrath, J. A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time? *Arch. Gen. Psychiatry* **64,** 1123–1131 (2007).
2. World Health Organization. *The Global Burden of Disease: 2004 Update* (WHO Press, 2008).
3. Knapp, M., Mangalore, R. & Simon, J. The global costs of schizophrenia. *Schizophr. Bull.* **30,** 279–293 (2004).
4. Lieberman, J. A. *et al.* Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N. Engl. J. Med.* **353,** 1209–1223 (2005).
5. Carlsson, A. & Lindqvist, M. Effect of chlorpromazine or haloperidol on formation of 3-methoxytyramine and normetanephrine in mouse brain. *Acta Pharmacol. Toxicol.* **20,** 140–144 (1963).
6. van Rossum, J. M. The significance of dopamine-receptor blockade for the mechanism of action of neuroleptic drugs. *Arch. Int. Pharmacodyn. Ther.* **160,** 492–494 (1966).
7. Lichtenstein, P. *et al.* Recurrence risks for schizophrenia in a Swedish national cohort. *Psychol. Med.* **36,** 1417–1425 (2006).
8. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60,** 1187–1192 (2003).
9. Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Rev. Genet.* **13,** 537–551 (2012).
10. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460,** 748–752 (2009).

11. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genet.* **45,** 1150–1159 (2013).

12. Ikeda, M. *et al.* Genome-wide association study of schizophrenia in a Japanese population. *Biol. Psychiatry* **69,** 472–478 (2011).

13. Hamshere, M. L. *et al.* Genome-wide significant associations in schizophrenia to *ITIH3/4, CACNA1C* and *SDCCAG8,* and extensive replication of associations reported by the Schizophrenia PGC. *Mol. Psychiatry* **18,** 708–712 (2013).

14. O'Donovan, M. C. *et al.* Identification of novel schizophrenia loci by genome-wide association and follow-up. *Nature Genet.* **40,** 1053–1055 (2008).

15. Rietschel, M. *et al.* Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. *Mol. Psychiatry* **17,** 906–917 (2012).

16. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nature Genet.* **43,** 969–976 (2011).

17. Irish Schizophrenia Genomics Consortium & Wellcome Trust Case Control Consortium. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biol. Psychiatry* **72,** 620–628 (2012).

18. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460,** 753–757 (2009).

19. Shi, Y. *et al.* Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nature Genet.* **43,** 1224–1227 (2011).

20. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460,** 744–747 (2009).

21. Steinberg, S. *et al.* Common variants at *VRK2* and *TCF4* conferring risk of schizophrenia. *Hum. Mol. Genet.* **20,** 4076–4081 (2011).

22. Yue, W. H. *et al.* Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nature Genet.* **43,** 1228–1231 (2011).

23. Lencz, T. *et al.* Genome-wide association study implicates *NDST3* in schizophrenia and bipolar disorder. *Nature Commun.* **4,** 2739 (2013).

24. Psychiatric GWAS Consortium. A framework for interpreting genomewide association studies of psychiatric disorders. *Mol. Psychiatry* **14,** 10–17 (2009).

25. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).

26. Begum, F., Ghosh, D., Tseng, G. C. & Feingold, E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res.* **40,** 3777–3784 (2012).

27. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).

28. Jostins, L. *et al.* Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491,** 119–124 (2012).

29. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19,** 807–812 (2011).

30. Bulik-Sullivan, B. K. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Preprint at http://dx.doi.org/10.1101/002931 (2014).

31. Ferreira, M. A. *et al.* Collaborative genome-wide association supports a role for *ANK3* and *CACNA1C* in bipolar disorder. *Nature Genet.* **40,** 1056–1058 (2008).

32. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381,** 1371–1379 (2013).

33. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506,** 185–190 (2014).

34. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506,** 179–184 (2014).

35. Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17,** 142–153 (2012).

36. Wellcome Trust Case Control Consortium Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genet.* **44,** 1294–1301 (2012).

37. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6,** e1000888 (2010).

38. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337,** 1190–1195 (2012).

39. Richards, A. L. *et al.* Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol. Psychiatry* **17,** 193–201 (2012).

40. Wright, F. A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nature Genet.* **46,** 430–437 (2014).

41. Doyle, J. P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135,** 749–762 (2008).

42. Tkachev, D. *et al.* Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet* **362,** 798–805 (2003).

43. Benros, M. E., Mortensen, P. B. & Eaton, W. W. Autoimmune diseases and infections as risk factors for schizophrenia. *Ann. NY Acad. Sci.* **1262,** 56–66 (2012).

44. Holmans, P. *et al.* Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am. J. Hum. Genet.* **85,** 13–24 (2009).

45. Lee, P. H., O'Dushlaine, C., Thomas, B. & Purcell, S. InRich: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* **28,** 1797–1799 (2012).

46. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36,** 214–224 (2012).

47. Gottesman, I. I. & Gould, T. D. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry* **160,** 636–645 (2003).

48. Insel, T. *et al.* Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167,** 748–751 (2010).

## Schizophrenia Working Group of the Psychiatric Genomics Consortium

Stephan Ripke[1,2], Benjamin M. Neale[1,2,3,4], Aiden Corvin[5], James T. R. Walters[6], Kai-How Farh[1], Peter A. Holmans[6,7], Phil Lee[1,2,4], Brendan Bulik-Sullivan[1,2], David A. Collier[8,9], Hailiang Huang[1,3], Tune H. Pers[3,10,11], Ingrid Agartz[12,13,14], Esben Agerbo[15,16,17], Margot Albus[18], Madeline Alexander[19], Farooq Amin[20,21], Silviu A. Bacanu[22], Martin Begemann[23], Richard A. Belliveau Jr[2], Judit Bene[24,25], Sarah E. Bergen[2,26], Elizabeth Bevilacqua[2], Tim B. Bigdeli[22], Donald W. Black[27], Richard Bruggeman[28], Nancy G. Buccola[29], Randy L. Buckner[30,31,32], William Byerley[33], Wiepke Cahn[34], Guiqing Cai[35,36], Dominique Campion[37], Rita M. Cantor[38], Vaughan J. Carr[39,40], Noa Carrera[6], Stanley V. Catts[39,41], Kimberly D. Chambert[2], Raymond C. K. Chan[42], Ronald Y. L. Chen[43], Eric Y. H. Chen[43,44], Wei Cheng[45], Eric F. C. Cheung[46], Siow Ann Chong[47], C. Robert Cloninger[48], David Cohen[49], Nadine Cohen[50], Paul Cormican[5], Nick Craddock[6,7], James J. Crowley[51], David Curtis[52,53], Michael Davidson[54], Kenneth L. Davis[36], Franziska Degenhardt[55,56], Jurgen Del Favero[57], Ditte Demontis[17,58,59], Dimitris Dikeos[60], Timothy Dinan[61], Srdjan Djurovic[14,62], Gary Donohoe[5,63], Elodie Drapeau[36], Jubao Duan[64,65], Frank Dudbridge[66], Naser Durmishi[67], Peter Eichhammer[68], Johan Eriksson[69,70,71], Valentina Escott-Price[6], Laurent Essioux[72], Ayman H. Fanous[73,74,75,76], Martilias S. Farrell[51], Josef Frank[77], Lude Franke[78], Robert Freedman[79], Nelson B. Freimer[80], Marion Friedl[81], Joseph I. Friedman[36], Menachem Fromer[1,2,4,82], Giulio Genovese[2], Lyudmila Georgieva[6], Ina Giegling[81,83], Paola Giusti-Rodríguez[51], Stephanie Godard[84], Jacqueline I. Goldstein[1,3], Vera Golimbet[85], Srihari Gopal[86], Jacob Gratten[87], Lieuwe de Haan[88], Christian Hammer[23], Marian L. Hamshere[6], Mark Hansen[89], Thomas Hansen[17,90], Vahram Haroutunian[36,91,92], Annette M. Hartmann[81], Frans A. Henskens[39,93,94], Stefan Herms[55,56,95], Joel N. Hirschhorn[3,11,96], Per Hoffmann[55,56,95], Andrea Hofman[55,56], Mads V. Hollegaard[97], David M. Hougaard[97], Masashi Ikeda[98], Inge Joa[99], Antonio Julià[100], René S. Kahn[34], Luba Kalaydjieva[101,102], Sena Karachanak-Yankova[103], Juha Karjalainen[78], David Kavanagh[6], Matthew C. Keller[104], James L. Kennedy[105,106,107], Andrey Khrunin[108], Yunjung Kim[51], Janis Klovins[109], James A. Knowles[110], Bettina Konte[81], Vaidutis Kucinskas[111], Zita Ausrele Kucinskiene[111], Hana Kuzelova-Ptackova[112], Anna K. Kähler[26], Claudine Laurent[19,113], Jimmy Lee Chee Keong[47,114], S. Hong Lee[87], Sophie E. Legge[6], Bernard Lerer[115], Miaoxin Li[43,44,116], Tao Li[117], Kung-Yee Liang[118], Jeffrey Lieberman[119], Svetlana Limborska[108], Carmel M. Loughland[39,120], Jan Lubinski[121], Jouko Lönnqvist[122], Milan Macek Jr[112], Patrik K. E. Magnusson[26], Brion S. Maher[123], Wolfgang Maier[124], Jacques Mallet[125], Sara Marsal[100], Manuel Mattheisen[17,58,59,126], Morten Mattingsdal[14,127], Robert W. McCarley[128,129], Colm McDonald[130], Andrew M. McIntosh[131,132], Sandra Meier[77], Carin J. Meijer[88], Bela Melegh[24,25], Ingrid Melle[14,133], Raquelle I. Mesholam-Gately[128,134], Andres Metspalu[135], Patricia T. Michie[39,136], Lili Milani[135], Vihra Milanova[137], Younes Mokrab[8], Derek W. Morris[5,63], Ole Mors[17,58,138], Kieran C. Murphy[139], Robin M. Murray[140], Inez Myin-Germeys[141], Bertram Müller-Myhsok[142,143,144], Mari Nelis[135], Igor Nenadic[145], Deborah A. Nertney[146], Gerald Nestadt[147], Kristin K. Nicodemus[148], Liene Nikitina-Zake[109], Laura Nisenbaum[149], Annelie Nordin[150], Eadbhard O'Callaghan[151], Colm O'Dushlaine[2], F. Anthony O'Neill[152], Sang-Yun Oh[153], Ann Olincy[79], Line Olsen[17,90], Jim Van Os[141,154], Psychosis Endophenotypes International Consortium[155], Christos Pantelis[39,156],

George N. Papadimitriou[60], Sergi Papiol[23], Elena Parkhomenko[36], Michele T. Pato[110], Tiina Paunio[157,158], Milica Pejovic-Milovancevic[159], Diana O. Perkins[160], Olli Pietiläinen[158,161], Jonathan Pimm[53], Andrew J. Pocklington[6], John Powell[140], Alkes Price[3,162], Ann E. Pulver[147], Shaun M. Purcell[82], Digby Quested[163], Henrik B. Rasmussen[17,90], Abraham Reichenberg[36], Mark A. Reimers[164], Alexander L. Richards[6], Joshua L. Roffman[30,32], Panos Roussos[82,165], Douglas M. Ruderfer[6,82], Veikko Salomaa[71], Alan R. Sanders[64,65], Ulrich Schall[39,120], Christian R. Schubert[166], Thomas G. Schulze[77,167], Sibylle G. Schwab[168], Edward M. Scolnick[2], Rodney J. Scott[39,169,170], Larry J. Seidman[128,134], Jianxin Shi[171], Engilbert Sigurdsson[172], Teimuraz Silagadze[173], Jeremy M. Silverman[36,174], Kang Sim[47], Petr Slominsky[108], Jordan W. Smoller[2,4], Hon-Cheong So[43], Chris C. A. Spencer[175], Eli A. Stahl[3,82], Hreinn Stefansson[176], Stacy Steinberg[176], Elisabeth Stogmann[177], Richard E. Straub[178], Eric Strengman[179,34], Jana Strohmaier[77], T. Scott Stroup[119], Mythily Subramaniam[47], Jaana Suvisaari[122], Dragan M. Svrakic[48], Jin P. Szatkiewicz[51], Erik Söderman[12], Srinivas Thirumalai[180], Draga Toncheva[103], Sarah Tosato[181], Juha Veijola[182,183], John Waddington[184], Dermot Walsh[185], Dai Wang[86], Qiang Wang[117], Bradley T. Webb[22], Mark Weiser[54], Dieter B. Wildenauer[186], Nigel M. Williams[6], Stephanie Williams[51], Stephanie H. Witt[77], Aaron R. Wolen[164], Emily H. M. Wong[43], Brandon K. Wormley[22], Hualin Simon Xi[187], Clement C. Zai[105,106], Xuebin Zheng[188], Fritz Zimprich[177], Naomi R. Wray[87], Kari Stefansson[176], Peter M. Visscher[87], Wellcome Trust Case-Control Consortium 2[189], Rolf Adolfsson[150], Ole A. Andreassen[14,133], Douglas H. R. Blackwood[132], Elvira Bramon[190], Joseph D. Buxbaum[35,36,91,191], Anders D. Børglum[17,58,59,138], Sven Cichon[55,56,95,192], Ariel Darvasi[193], Enrico Domenici[194], Hannelore Ehrenreich[23], Tõnu Esko[3,11,96,135], Pablo V. Gejman[64,65], Michael Gill[5], Hugh Gurling[53], Christina M. Hultman[26], Nakao Iwata[98], Assen V. Jablensky[39,102,186,195], Erik G. Jönsson[12,14], Kenneth S. Kendler[196], George Kirov[6], Jo Knight[105,106,107], Todd Lencz[197,198,199], Douglas F. Levinson[19], Qingqin S. Li[86], Jianjun Liu[188,200], Anil K. Malhotra[197,198,199], Steven A. McCarroll[2,96], Andrew McQuillin[53], Jennifer L. Moran[2], Preben B. Mortensen[15,16,17], Bryan J. Mowry[87,201], Markus M. Nöthen[55,56], Roel A. Ophoff[38,80,34], Michael J. Owen[6,7], Aarno Palotie[2,4,161], Carlos N. Pato[110], Tracey L. Petryshen[2,128,202], Danielle Posthuma[203,204,205], Marcella Rietschel[77], Brien P. Riley[196], Dan Rujescu[81,83], Pak C. Sham[43,44,116], Pamela Sklar[82,91,165], David St Clair[206], Daniel R. Weinberger[178,207], Jens R. Wendland[166], Thomas Werge[17,90,208], Mark J. Daly[1,2,3], Patrick F. Sullivan[26,51,160] & Michael C. O'Donovan[6,7]

[1]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [3]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [4]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [5]Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Dublin 8, Ireland. [6]MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff CF24 4HQ, UK. [7]National Centre for Mental Health, Cardiff University, Cardiff CF24 4HQ, UK. [8]Eli Lilly and Company Limited, Erl Wood Manor, Sunninghill Road, Windlesham, Surrey GU20 6PH, UK. [9]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London SE5 8AF, UK. [10]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, DK-2800, Denmark. [11]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts 02115, USA. [12]Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, SE-17176 Stockholm, Sweden. [13]Department of Psychiatry, Diakonhjemmet Hospital, 0319 Oslo, Norway. [14]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, 0424 Oslo, Norway. [15]Centre for Integrative Register-based Research, CIRRAU, Aarhus University, DK-8210 Aarhus, Denmark. [16]National Centre for Register-based Research, Aarhus University, DK-8210 Aarhus, Denmark. [17]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark. [18]State Mental Hospital, 85540 Haar, Germany. [19]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305, USA. [20]Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Atlanta, Georgia 30033, USA. [21]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia 30322, USA. [22]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia 23298, USA. [23]Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Göttingen 37075, Germany. [24]Department of Medical Genetics, University of Pécs, Pécs H-7624, Hungary. [25]Szentagothai Research Center, University of Pécs, Pécs H-7624, Hungary. [26]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-17177, Sweden. [27]Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa 52242, USA. [28]University Medical Center Groningen, Department of Psychiatry, University of Groningen NL-9700 RB, The Netherlands. [29]School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana 70112, USA. [30]Athinoula A. Martinos Center, Massachusetts General Hospital, Boston, Massachusetts 02129, USA. [31]Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA. [32]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [33]Department of Psychiatry, University of California at San Francisco, San Francisco, California 94143, USA. [34]University Medical Center Utrecht, Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, 3584 Utrecht, The Netherlands. [35]Department of Human Genetics, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [36]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [37]Centre Hospitalier du Rouvray and INSERM U1079 Faculty of Medicine, 76301 Rouen, France. [38]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA. [39]Schizophrenia Research Institute, Sydney NSW 2010, Australia. [40]School of Psychiatry, University of New South Wales, Sydney NSW 2031, Australia. [41]Royal Brisbane and Women's Hospital, University of Queensland, Brisbane, St Lucia QLD 4072, Australia. [42]Institute of Psychology, Chinese Academy of Science, Beijing 100101, China. [43]Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [44]State Key Laboratory for Brain and Cognitive Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [45]Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina 27514, USA. [46]Castle Peak Hospital, Hong Kong, China. [47]Institute of Mental Health, Singapore 539747, Singapore. [48]Department of Psychiatry, Washington University, St. Louis, Missouri 63110, USA. [49]Department of Child and Adolescent Psychiatry, Assistance Publique Hopitaux de Paris, Pierre and Marie Curie Faculty of Medicine and Institute for Intelligent Systems and Robotics, Paris 75013, France. [50]Blue Note Biosciences, Princeton, New Jersey 08540, USA [51]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599-7264, USA. [52]Department of Psychological Medicine, Queen Mary University of London, London E1 1BB, UK. [53]Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London WC1E 6JJ, UK. [54]Sheba Medical Center, Tel Hashomer 52621, Israel. [55]Department of Genomics, Life and Brain Center, D-53127 Bonn, Germany. [56]Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany. [57]Applied Molecular Genomics Unit, VIB Department of Molecular Genetics, University of Antwerp, B-2610 Antwerp, Belgium. [58]Centre for Integrative Sequencing, iSEQ, Aarhus University, DK-8000 Aarhus C, Denmark. [59]Department of Biomedicine, Aarhus University, DK-8000 Aarhus C, Denmark. [60]First Department of Psychiatry, University of Athens Medical School, Athens 11528, Greece. [61]Department of Psychiatry, University College Cork, Co. Cork, Ireland. [62]Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway. [63]Cognitive Genetics and Therapy Group, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Co. Galway, Ireland. [64]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois 60637, USA. [65]Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, Illinois 60201, USA. [66]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. [67]Department of Child and Adolescent Psychiatry, University Clinic of Psychiatry, Skopje 1000, Republic of Macedonia. [68]Department of Psychiatry, University of Regensburg, 93053 Regensburg, Germany. [69]Department of General Practice, Helsinki University Central Hospital, University of Helsinki P.O. Box 20, Tukholmankatu 8 B, FI-00014, Helsinki, Finland [70]Folkhälsan Research Center, Helsinki, Finland, Biomedicum Helsinki 1, Haartmaninkatu 8, FI-00290, Helsinki, Finland. [71]National Institute for Health and Welfare, P.O. Box 30, FI-00271 Helsinki, Finland. [72]Translational Technologies and Bioinformatics, Pharma Research and Early Development, F. Hoffman-La Roche, CH-4070 Basel, Switzerland. [73]Department of Psychiatry, Georgetown University School of Medicine, Washington DC 20057, USA. [74]Department of Psychiatry, Keck School of Medicine of the University of Southern California, Los Angeles, California 90033, USA. [75]Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, Virginia 23298, USA. [76]Mental Health Service Line, Washington VA Medical Center, Washington DC 20422, USA. [77]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg , D-68159 Mannheim, Germany. [78]Department of Genetics, University of Groningen, University Medical Centre Groningen, 9700 RB Groningen, The Netherlands. [79]Department of Psychiatry, University of Colorado Denver, Aurora, Colorado 80045, USA. [80]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California 90095, USA. [81]Department of Psychiatry, University of Halle, 06112 Halle, Germany. [82]Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, New York 10029, USA. [83]Department of Psychiatry, University of Munich, 80336, Munich, Germany. [84]Departments of Psychiatry and Human and Molecular Genetics, INSERM, Institut de Myologie, Hôpital de la Pitiè-Salpêtrière, Paris 75013, France. [85]Mental Health Research Centre, Russian Academy of Medical Sciences, 115522 Moscow, Russia. [86]Neuroscience Therapeutic Area, Janssen Research and Development, Raritan, New Jersey 08869, USA. [87]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, QLD 4072, Australia. [88]Academic Medical Centre University of Amsterdam, Department of Psychiatry, 1105 AZ Amsterdam, The Netherlands. [89]Illumina, La Jolla, California, California 92122, USA. [90]Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Mental Health Services Copenhagen, DK-4000, Denmark. [91]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [92]J. J. Peters VA Medical Center, Bronx, New York, New York 10468, USA. [93]Priority Research Centre for Health Behaviour, University of Newcastle, Newcastle NSW 2308, Australia. [94]School of Electrical Engineering and Computer Science, University of Newcastle, Newcastle NSW 2308, Australia. [95]Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel CH-4058, Switzerland. [96]Department of Genetics, Harvard Medical School, Boston, Massachusetts, Massachusetts 02115, USA. [97]Section of Neonatal Screening and Hormones, Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum Institut, Copenhagen DK-2300, Denmark. [98]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Aichi,470-1192, Japan. [99]Regional Centre for Clinical Research in Psychosis, Department of Psychiatry, Stavanger University Hospital, 4011 Stavanger, Norway. [100]Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona 08035, Spain. [101]Centre for Medical Research, The University of Western Australia, Perth WA6009, Australia. [102]The Perkins Institute for Medical Research, The University of Western Australia, Perth WA6009, Australia. [103]Department of Medical Genetics, Medical University, Sofia 1431, Bulgaria. [104]Department of Psychology, University of Colorado Boulder, Boulder, Colorado 80309, USA. [105]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario M5T 1R8, Canada. [106]Department of Psychiatry, University of Toronto, Toronto, Ontario M5T 1R8, Canada. [107]Institute of Medical Science, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [108]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia. [109]Latvian Biomedical Research and Study Centre, Riga, LV-1067, Latvia. [110]Department of Psychiatry and Zilkha Neurogenetics Institute, Keck School of Medicine at University of

Southern California, Los Angeles, California 90089, USA. [111]Faculty of Medicine, Vilnius University, LT-01513 Vilnius, Lithuania. [112]Department of Biology and Medical Genetics, 2nd Faculty of Medicine and University Hospital Motol, 150 06 Prague, Czech Republic. [113]Department of Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Paris 75013, France. [114]Duke-NUS Graduate Medical School, Singapore 169857. [115]Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel. [116]Centre for Genomic Sciences, The University of Hong Kong, Hong Kong, China. [117]Mental Health Centre and Psychiatric Laboratory, West China Hospital, Sichuan University, Chengdu, 610041 Sichuan, China. [118]Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland 21205, USA. [119]Department of Psychiatry, Columbia University, New York, New York 10032, USA. [120]Priority Centre for Translational Neuroscience and Mental Health, University of Newcastle, Newcastle NSW 2300, Australia. [121]Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University in Szczecin, 70-453 Szczecin, Poland. [122]Department of Mental Health and Substance Abuse Services; National Institute for Health and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland. [123]Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA. [124]Department of Psychiatry, University of Bonn, D-53127 Bonn, Germany. [125]Centre National de la Recherche Scientifique, Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, Hôpital de la Pitié Salpêtrière, 75013 Paris, France. [126]Department of Genomics Mathematics, University of Bonn, D-53127 Bonn, Germany. [127]Research Unit, Sørlandet Hospital, 4604 Kristiansand, Norway. [128]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA. [129]VA Boston Health Care System, Brockton, Massachusetts 02301, USA. [130]Department of Psychiatry, National University of Ireland Galway, Co. Galway, Ireland. [131]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH16 4SB, UK. [132]Division of Psychiatry, University of Edinburgh, Edinburgh EH16 4SB, UK. [133]Division of Mental Health and Addiction, Oslo University Hospital, 0424 Oslo, Norway. [134]Massachusetts Mental Health Center Public Psychiatry Division of the Beth Israel Deaconess Medical Center, Boston, Massachusetts 02114, USA. [135]Estonian Genome Center, University of Tartu, Tartu 50090, Estonia. [136]School of Psychology, University of Newcastle, Newcastle NSW 2308, Australia. [137]First Psychiatric Clinic, Medical University, Sofia 1431, Bulgaria. [138]Department P, Aarhus University Hospital, DK-8240 Risskov, Denmark. [139]Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin 2, Ireland. [140]King's College London, London SE5 8AF, UK. [141]Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, EURON, 6229 HX Maastricht, The Netherlands. [142]Institute of Translational Medicine, University of Liverpool, Liverpool L69 3BX, UK. [143]Max Planck Institute of Psychiatry, 80336 Munich, Germany. [144]Munich Cluster for Systems Neurology (SyNergy), 80336 Munich, Germany. [145]Department of Psychiatry and Psychotherapy, Jena University Hospital, 07743 Jena, Germany. [146]Department of Psychiatry, Queensland Brain Institute and Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, St Lucia QLD 4072, Australia. [147]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [148]Department of Psychiatry, Trinity College Dublin, Dublin 2, Ireland. [149]Eli Lilly and Company, Lilly Corporate Center, Indianapolis, 46285 Indiana, USA. [150]Department of Clinical Sciences, Psychiatry, Umeå University, SE-901 87 Umeå, Sweden. [151]DETECT Early Intervention Service for Psychosis, Blackrock, Co. Dublin, Ireland. [152]Centre for Public Health, Institute of Clinical Sciences, Queen's University Belfast, Belfast BT12 6AB, UK. [153]Lawrence Berkeley National Laboratory, University of California at Berkeley, Berkeley, California 94720, USA. [154]Institute of Psychiatry, King's College London, London SE5 8AF, UK. [155]A list of authors and affiliations appear in the Supplementary Information. [156]Melbourne Neuropsychiatry Centre, University of Melbourne & Melbourne Health, Melbourne, Vic 3053, Australia. [157]Department of Psychiatry, University of Helsinki, P.O. Box 590, FI-00029 HUS, Helsinki, Finland. [158]Public Health Genomics Unit, National Institute for Health and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland [159]Medical Faculty, University of Belgrade, 11000 Belgrade, Serbia. [160]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27599-7160, USA. [161]Institute for Molecular Medicine Finland, FIMM, University of Helsinki, P.O. Box 20FI-00014, Helsinki, Finland [162]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [163]Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, UK. [164]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia 23298, USA. [165]Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [166]PharmaTherapeutics Clinical Research, Pfizer Worldwide Research and Development, Cambridge, Massachusetts 02139, USA. [167]Department of Psychiatry and Psychotherapy, University of Gottingen, 37073 Göttingen, Germany. [168]Psychiatry and Psychotherapy Clinic, University of Erlangen, 91054 Erlangen, Germany. [169]Hunter New England Health Service, Newcastle NSW 2308, Australia. [170]School of Biomedical Sciences, University of Newcastle, Newcastle NSW 2308, Australia. [171]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA. [172]University of Iceland, Landspitali, National University Hospital, 101 Reykjavik, Iceland. [173]Department of Psychiatry and Drug Addiction, Tbilisi State Medical University (TSMU), N33, 0177 Tbilisi, Georgia. [174]Research and Development, Bronx Veterans Affairs Medical Center, New York, New York 10468, USA. [175]Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK. [176]deCODE Genetics, 101 Reykjavik, Iceland. [177]Department of Clinical Neurology, Medical University of Vienna, 1090 Wien, Austria. [178]Lieber Institute for Brain Development, Baltimore, Maryland 21205, USA. [179]Department of Medical Genetics, University Medical Centre Utrecht, Universiteitsweg 100, 3584 CG, Utrecht, The Netherlands. [180]Berkshire Healthcare NHS Foundation Trust, Bracknell RG12 1BQ, UK. [181]Section of Psychiatry, University of Verona, 37134 Verona, Italy. [182]Department of Psychiatry, University of Oulu, P.O. Box 5000, 90014, Finland. [183]University Hospital of Oulu, P.O. Box 20, 90029 OYS, Finland. [184]Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland. [185]Health Research Board, Dublin 2, Ireland. [186]School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth WA6009, Australia. [187]Computational Sciences CoE, Pfizer Worldwide Research and Development, Cambridge, Massachusetts 02139, USA. [188]Human Genetics, Genome Institute of Singapore, A*STAR, Singapore 138672. [189]A list of authors and affiliations appear in the Supplementary Information. [190]University College London, London WC1E 6BT, UK. [191]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [192]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, 52428 Juelich, Germany. [193]Department of Genetics, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel. [194]Neuroscience Discovery and Translational Area, Pharma Research and Early Development, F. Hoffman-La Roche, CH-4070 Basel, Switzerland. [195]Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Medical Research Foundation Building, Perth WA6000, Australia. [196]Virginia Institute for Psychiatric and Behavioral Genetics, Departments of Psychiatry and Human and Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia 23298, USA. [197]The Feinstein Institute for Medical Research, Manhasset, New York 11030, USA. [198]The Hofstra NS-LIJ School of Medicine, Hempstead, New York 11549, USA. [199]The Zucker Hillside Hospital, Glen Oaks, New York 11004, USA. [200]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore. [201]Queensland Centre for Mental Health Research, University of Queensland, Brisbane 4076, Queensland, Australia. [202]Center for Human Genetic Research and Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [203]Department of Child and Adolescent Psychiatry, Erasmus University Medical Centre, Rotterdam 3000, The Netherlands. [204]Department of Complex Trait Genetics, Neuroscience Campus Amsterdam, VU University Medical Center Amsterdam, Amsterdam 1081, The Netherlands. [205]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University, Amsterdam 1081, The Netherlands. [206]University of Aberdeen, Institute of Medical Sciences, Aberdeen AB25 2ZD, UK. [207]Departments of Psychiatry, Neurology, Neuroscience and Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA. [208]Department of Clinical Medicine, University of Copenhagen, Copenhagen 2200, Denmark.
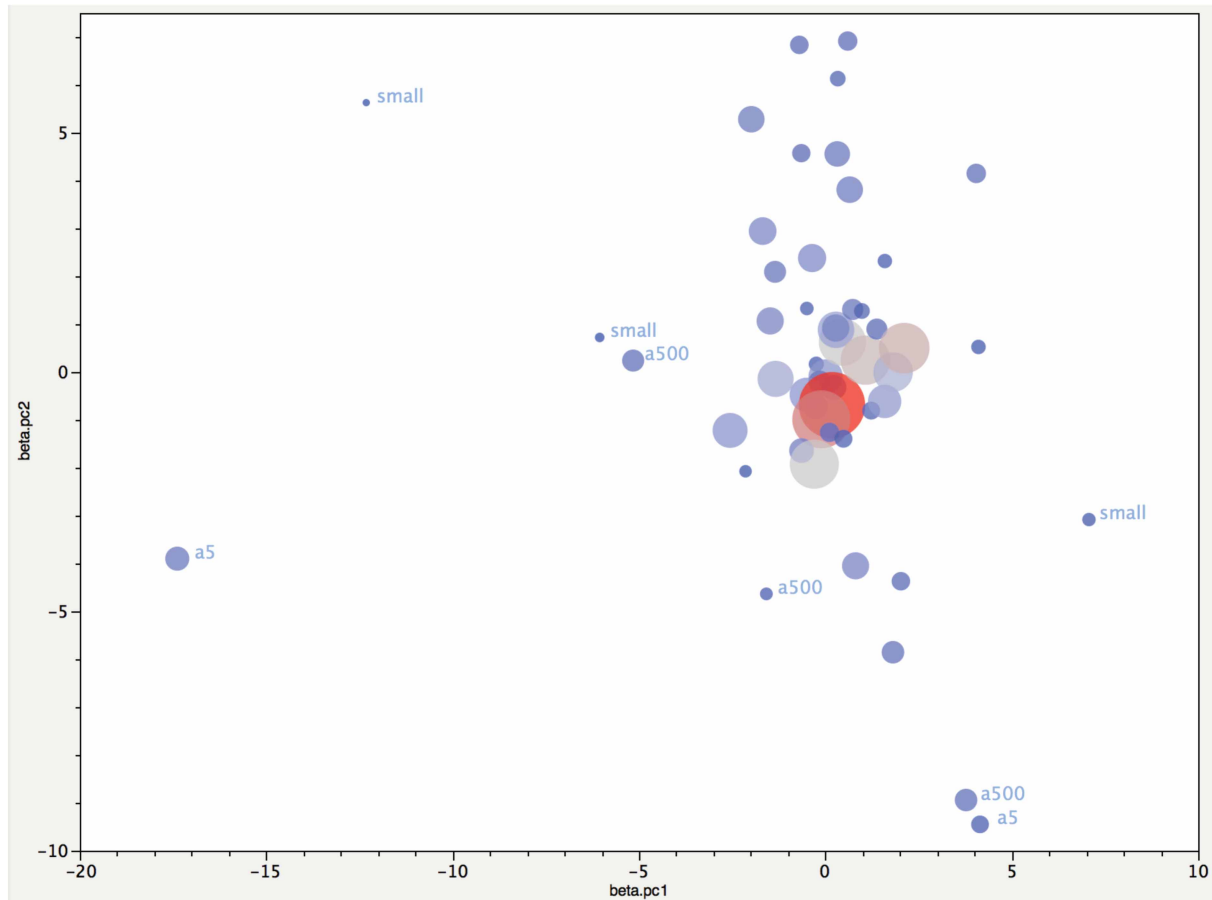
Extended Data Figure 1 | Homogeneity of effects across studies. Plot of the first two principal components (PCs) from principal components analysis (PCA) of the logistic regression β coefficients for autosomal genome-wide significant associations. The input data were the β coefficients from 52 samples for 112 independent SNP associations (excluding 3 chrX SNPs and 13 SNPs with missing values in Asian samples). PCAs were weighted by the number of cases. Each circle shows the location of a study on PC1 and PC2. Circle size and colour are proportional to the number of cases in each sample (larger and darker red circles correspond to more cases). Most samples cluster. Outliers had either small numbers of cases ('small') or were genotyped on older arrays. Abbreviations: a500 (Affymetrix 500K); a5 (Affymetrix 5.0). Studies that did not use conventional research interviews are in the central cluster (CLOZUK, Sweden, and Denmark-Aarhus studies, see Supplementary Methods for sample descriptions).
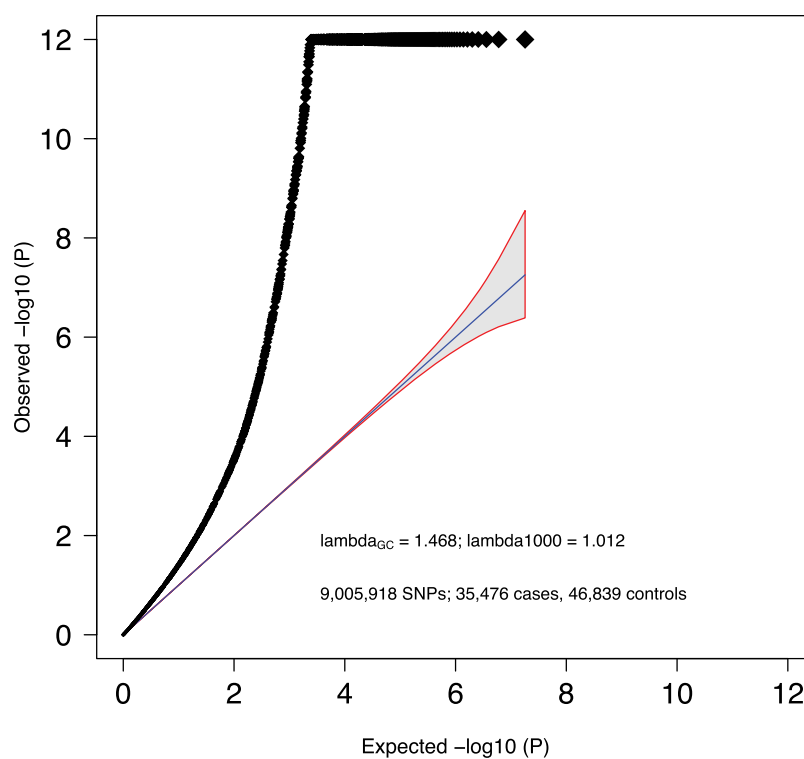
Extended Data Figure 2 | Quantile-quantile plot. Quantile-quantile plot of the discovery genome-wide association meta-analysis of 49 case control samples (34,241 cases and 45,604 controls) and 3 family based association studies (1,235 parent affected-offspring trios). Expected $-\log_{10} P$ values are those expected under the null hypothesis. Observed are the GWAS association results derived by logistic regression (2-tailed) as in Fig. 1. For clarity, we avoided expansion of the y axis by setting the smallest association P values to 10^{-12}. The shaded area surrounded by a red line indicates the 95% confidence interval under the null. λ_{GC} is the observed median χ^2 test statistic divided by the median expected χ^2 test statistic under the null hypothesis.

a


```
Lambda GC = 1.325
Mean Chi-Square = 1.314
Slope = -4.963e-05
Intercept = 1.319
```

b


```
Lambda GC = 1.324
Mean Chi-Square = 1.429
Slope = 0.004684
Intercept = 1.018
```

c


```
Lambda GC = 1.484
Mean Chi-Square = 1.613
Slope = 0.005209
Intercept = 1.066
```
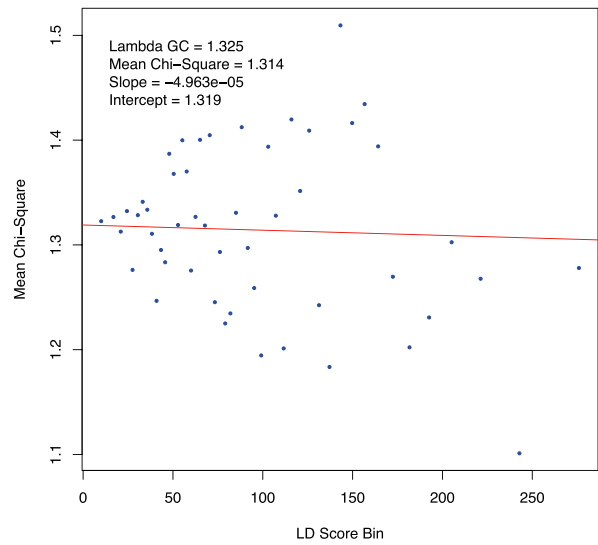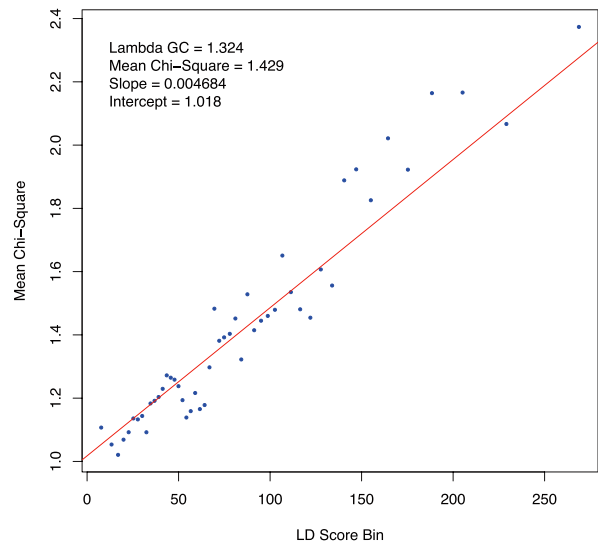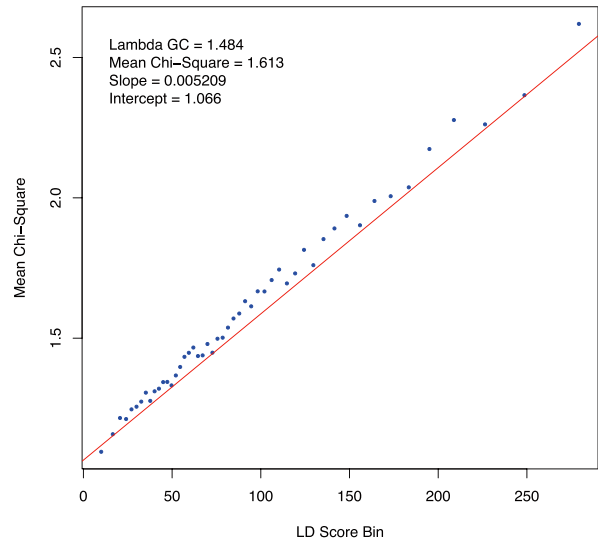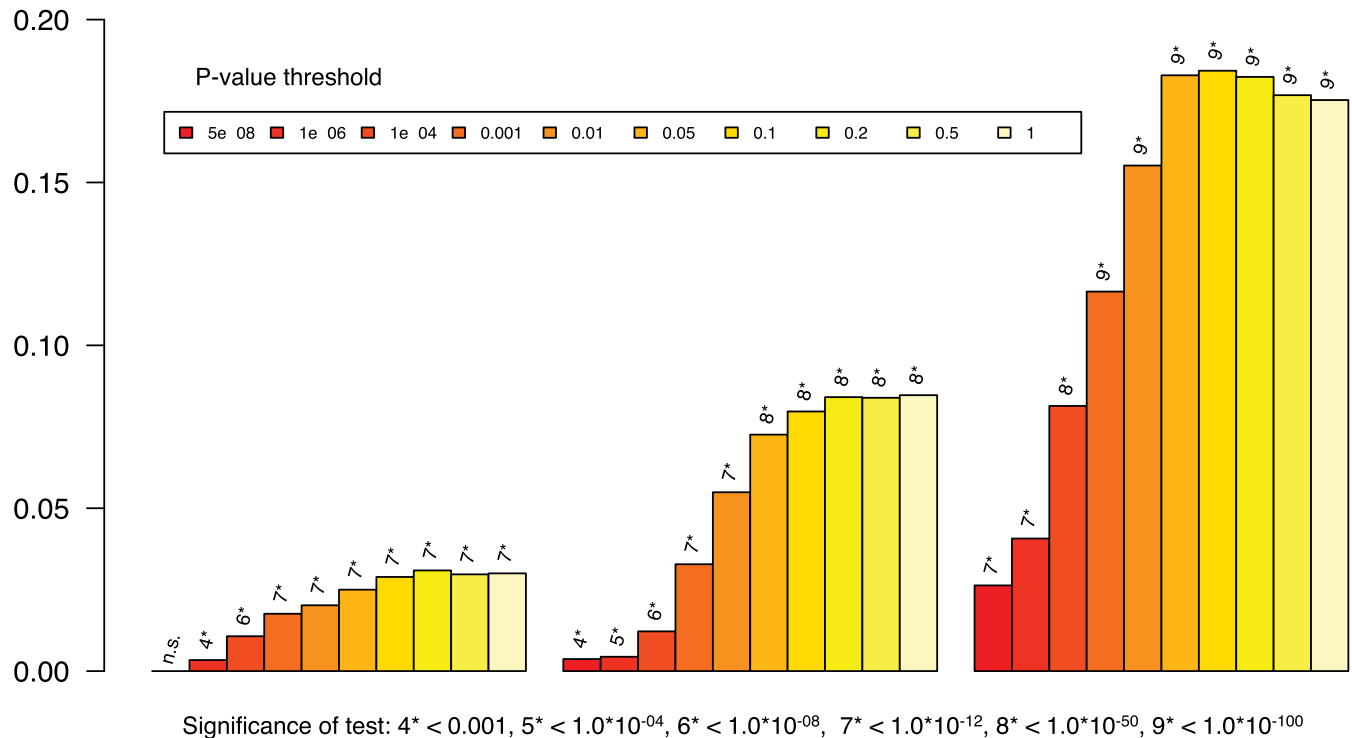
Extended Data Figure 3 | Linkage disequilibrium score regression consistent with polygenic inheritance. The relationship between marker χ^2 association statistics and linkage disequilibrium (LD) as measured by the linkage disequilibrium score. Linkage disequilibrium score is the sum of the r^2 values between a variant and all other known variants within a 1 cM window, and quantifies the amount of genetic variation tagged by that variant. Variants were grouped into 50 equal-sized bins based on linkage disequilibrium score rank. Linkage disequilibrium score bin and mean χ^2 denotes mean linkage disequilibrium score and test statistic for markers each bin. **a, b,** We simulated (Supplementary Methods) test statistics under two scenarios: **a,** no true association, inflation due to population stratification; and **b,** polygenic inheritance ($\lambda = 1.32$), in which we assigned independent and identically distributed per-normalized-genotype effects to a randomly selected subset of variants. **c,** Results from the PGC schizophrenia GWAS ($\lambda = 1.48$). The real data are strikingly similar to the simulated data summarized in **b** but not **a**. The intercept estimates the inflation in the mean χ^2 that results from confounding biases, such as cryptic relatedness or population stratification. Thus, the intercept of 1.066 for the schizophrenia GWAS suggests that ~90% of the inflation in the mean χ^2 results from polygenic signal. The results of the simulations are also consistent with theoretical expectation (see Supplementary Methods). λ is the median χ^2 test statistic from the simulations (**a, b**) or the observed data (**c**) divided by the median expected χ^2 test statistic under the null hypothesis.
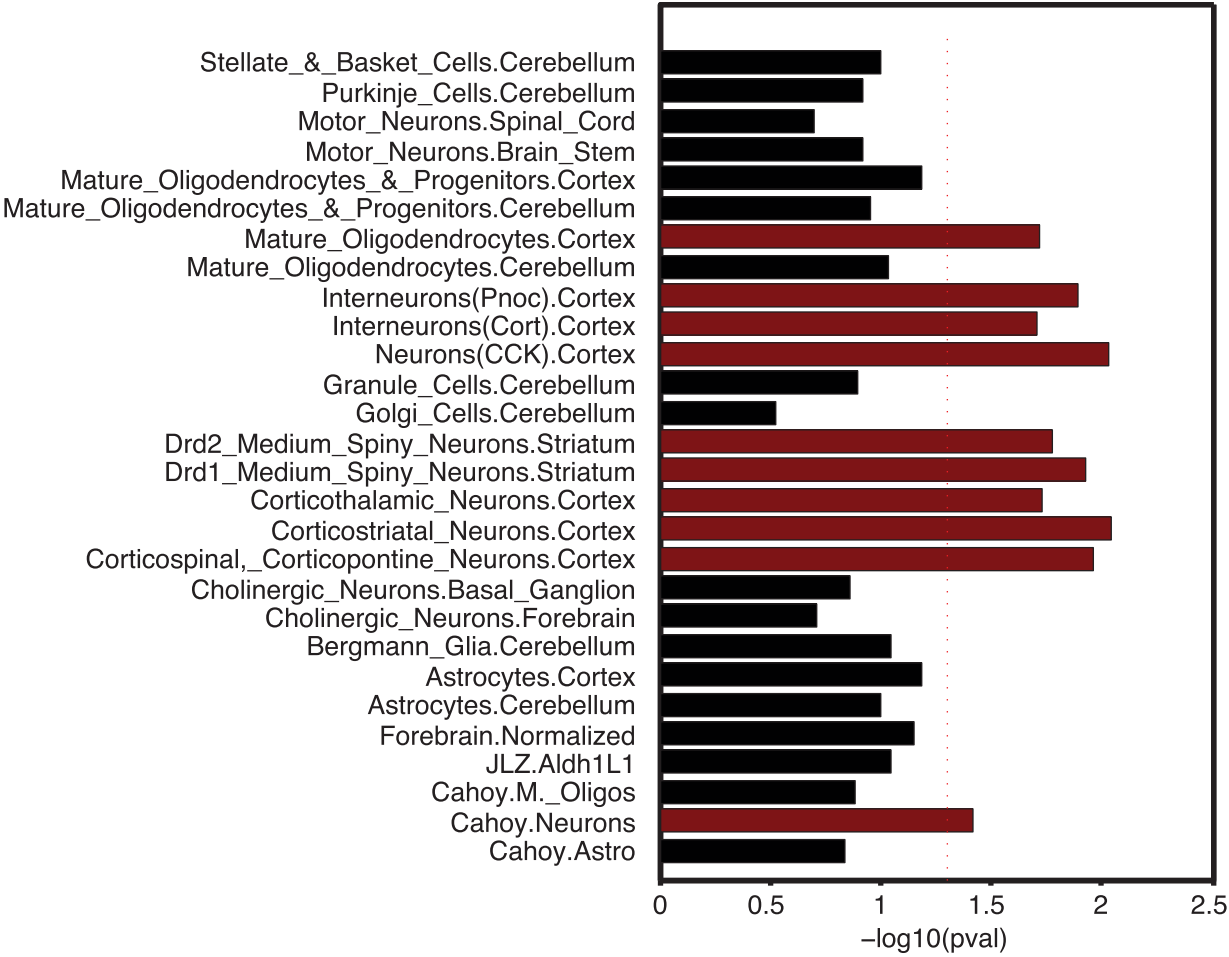
Extended Data Figure 4 | Enrichment of associations in tissues and cells. Genes whose transcriptional start is nearest to the most associated SNP at each schizophrenia-associated locus were tested for enriched expression in purified brain cell subsets obtained from mouse ribotagged lines[41] using enrichment analysis described in the Supplementary Methods. The red dotted line indicates $P = 0.05$.

Nagelkerke R^2

Significance of test: $4* < 0.001$, $5* < 1.0*10^{-04}$, $6* < 1.0*10^{-08}$, $7* < 1.0*10^{-12}$, $8* < 1.0*10^{-50}$, $9* < 1.0*10^{-100}$

Extended Data Figure 5 | MGS risk profile score analysis. Polygenic risk profile score (RPS) analyses using the MGS[18] sample as target, and deriving risk alleles from three published schizophrenia data sets (x axis): ISC (2,615 cases and 3,338 controls)[10], PGC1 (excluding MGS, 9,320 cases and 10,228 controls)[16], and the current meta-analysis (excluding MGS) with 32,838 cases and 44,357 controls. Samples sizes differ slightly from the original publications due to different analytical procedures. This shows the increasing RPS prediction with increasing training data set size reflecting improved precision of estimates of the SNP effect sizes. The proportion of variance explained (y axis; Nagelkerke's R^2) was computed by comparison of a full model (covariates + RPS) score to a reduced model (covariates only). Ten different P value thresholds (P_T) for selecting risk alleles are denoted by the colour of each bar (legend above plot). For significance testing, see the bottom legend which denotes the P value for the test that R^2 is different from zero. All numerical data and methods used to generate these plots are available in Supplementary Table 6 and Supplementary Methods.
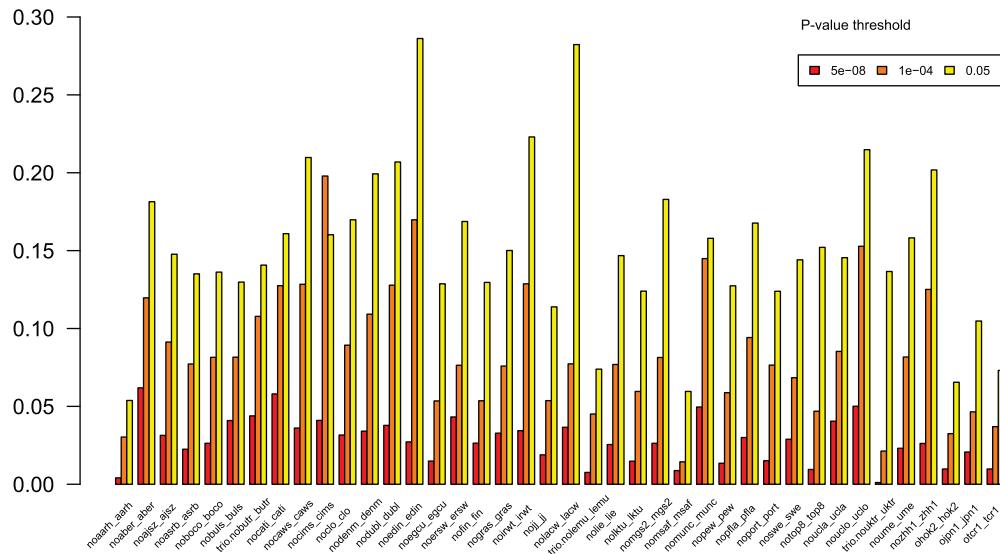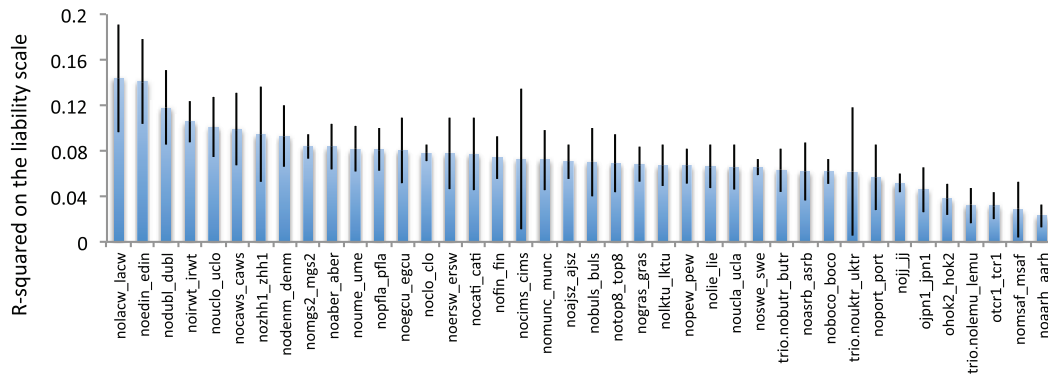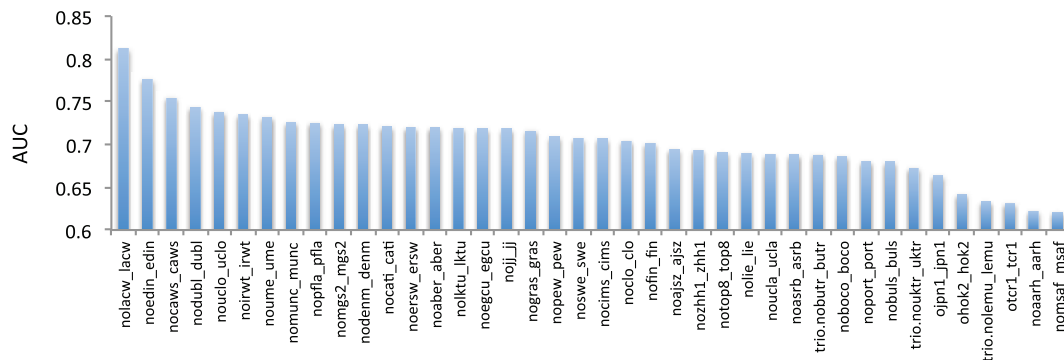
Extended Data Figure 6 | Risk profile score analysis. We defined 40 target subgroups of the primary GWAS data set and performed 40 leave-one-out GWAS analyses (see Supplementary Methods and Supplementary Table 7) from which we derived risk alleles for RPS analysis (x axis) for each target subgroup. **a**, The proportion of variance explained (y axis; Nagelkerke's R^2) was computed for each target by comparison of a full model (covariates + RPS) score to a reduced model (covariates only). For clarity, 3 different P value thresholds (P_T) are presented denoted by the colour of each bar (legend above plot) as for Extended Data Fig. 5, but for clarity we restrict to fewer P value thresholds (P_T of 5×10^{-8}, 1×10^{-4} and 0.05) and removed the significance values. **b**, The proportion of variance on the liability scale from risk scores calculated at the P_T 0.05 with 95% CI bar assuming baseline population disease risk of 1%. **c**, Area under the receiver operating curve (AUC). All numerical data and methods used to generate these plots are available in Supplementary Table 7 and Supplementary Methods.
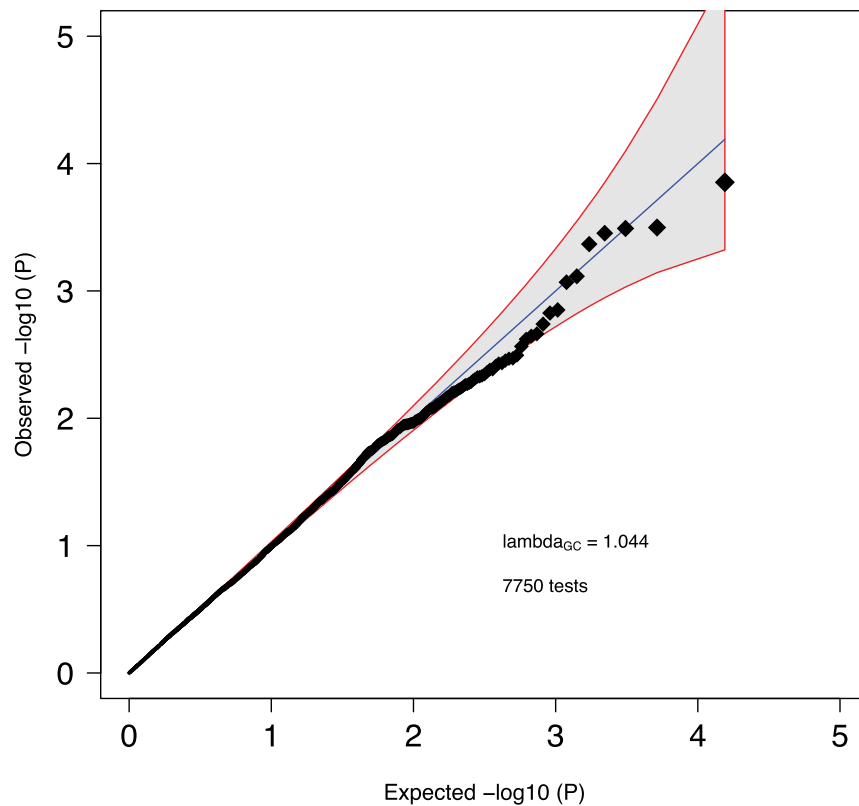
Extended Data Figure 7 | Pairwise epistasis analysis of significant SNPs.
Quantile-quantile plot for all pair-wise ($n = 7,750$) combinations of the 125 independent autosomal genome-wide significant SNPs tested for non-additive effects on risk using case-control data sets of European ancestry (32,405 cases and 42,221 controls). We included as covariates the principal components from the main analysis as well as a study indicator. The interaction model is described by:

$$Y = \beta_0 + \hat{a}_1 X_1 + \hat{a}_2 X_2 + \hat{a}_3 * X_1^* X_2 + \hat{a}_4 X_4 + \hat{a}_5 X_5$$

X_1 and X_2 are genotypes at the two loci, $X_1^* X_2$ is the interaction between the two genotypes modelled in a multiplicative fashion, X_4 is the vector of principal components, X_5 is the vector of study indicator variables. Each $\hat{a}$ is the regression coefficient in the generalized linear model using logistic regression. The overall distribution of P values did not deviate from the null and the smallest P value (4.28×10^{-4}) did not surpass the Bonferroni correction threshold ($P = 0.05/7750 = 6.45 \times 10^{-6}$). The line $x = y$ indicates the expected null distribution with the grey area bounded by red lines indicating the expected 95% confidence interval for the null.

Extended Data Table 1 | ALIGATOR and INRICH

SET	ALIGATOR	INRICH
Postsynaptic sets		
ARC	NA	1
NMDAR	NA	0.458
Curated pre- and postsynaptic sets		
Cell adhesion and trans-synaptic signalling	0.902	0.44
Structural plasticity	NA	NA
Excitability	NA	NA
FMRP sets		
FMRP	**0.0066**	$\mathbf{5 \times 10^{-5}}$
MIR137 sets		
Targetscan v5 with PCT > 0.9	**0.0371**	**0.0103**
Targetscan v6.2	0.059	**0.0024**
Calcium signalling sets		
CACN* channel subunits	**0.0338**	**0.022**

Gene sets that have been reported to be enriched for schizophrenia associations and or rare mutations were tested for enrichment for genome-wide significant associations using ALIGATOR[44] and INRICH[45]. Specifically, we tested the glutamatergic postsynaptic proteins comprising activity-regulated cytoskeleton-associated protein (ARC) and *N*-methyl-ᴅ-aspartate receptor (NMDAR) complexes[33–35], other curated synaptic gene-sets[14,49], targets of fragile X mental retardation protein (FMRP)[33–35], calcium channels[11,33], and TargetScan predicted MIR137 sets[11,16]. The MIR137 TargetScan sets contain computationally predicted conserved miRNA target sites in 3′ UTRs of human genes[50]. The current version is v6, but the version used in the prior PGC SCZ report[16] was based on v5 (filtered for a probability of conserved targeting > 0.9). We report the results of both analyses for consistency with previous work. The association at the extended MHC complex was not included given the extensive linkage disequilibrium at this region spans large numbers of genes. NA means that the pathway in question contained fewer than 2 significant genes (for ALIGATOR) or regions (INRICH).

49. Lips, E. S. *et al.* Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Mol. Psychiatry* **17,** 996–1006 (2012).

50. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).

Extended Data Table 2 | *de novo* overlap

Disease group	NS (N)	*P*	# NS in PGC2 loci	Observed (stat)	Expected (stat)	Genes
SCZ	702	**0.0061**	25	10.97	5.27	CACNA1I(x2) *CCDC39* CD14(x2) CR1L CUL3 DPEP2 *DPYD*(x2) EP300 *ESAM* GRIN2A *LRP1* NCAN PDCD11 PTPRF RIMS1 SBNO1 SGSM2 SLC7A6 STAG1 TMEM219 *ZDHHC5* ZNF536
ID	141	**0.00002**	11	6.87	1.05	GRIA1 GRIN2A(x2) LRP1 NEK1 NGEF SATB2 SREBF2 STAG1 TCF4(x2)
ASD	789	**0.035**	19	9.99	5.93	*APH1A* CNOT1 CSMD1 *CUL3* CYP17A1 CYP26B1 EPHX2 LRP1 MAPK3 MEF2C MPP6 MYO15A NISCH PBRM1 PRKD1 *RIMS1* TSNARE1 WDR55 ZNF804A
Controls	434	0.15	16	4.88	3.28	ANKRD44 C11orf87 CCDC39 CDK2AP1 CHRM4 DPEP2 EP300 LRP1 LRRC48 MAN2A1 MYO1A OSBPL3 RAI1 SF3B1 SREBF2 TLE3

Test of overlap between genes mapping to schizophrenia-associated loci in the present study and genes affected by non-synonymous (NS) *de novo* mutations. Enrichment was calculated using the dnenrich permutation framework as described[34]. Genes within the GWS loci (Supplementary Table 3) were weighted by $1/N$, where N is the number of coding genes within each associated locus. The observed test statistic (stat) is the sum of weights of genes impacted by *de novo* mutations. The expected test statistics are calculated by averaging over 50,000 permuted *de novo* mutation sets. Genes within schizophrenia-associated loci affected by *de novo* mutations are listed (multiple hits listed in parentheses). Cohorts: SCZ, schizophrenia; ID, intellectual disability; ASD, autism spectrum disorder. All mutations analysed annotated according to a unified system (see Supplementary Tables 1 and 2 of ref. 34). Genes with loss-of-function *de novo* mutations are underlined and in italics.

ARTICLE

Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma

Paul A. Northcott[1]*, Catherine Lee[2,3]*, Thomas Zichner[4]*, Adrian M. Stütz[4], Serap Erkek[1,4], Daisuke Kawauchi[1], David J. H. Shih[5], Volker Hovestadt[6], Marc Zapatka[6], Dominik Sturm[1], David T. W. Jones[1], Marcel Kool[1], Marc Remke[5], Florence M. G. Cavalli[5], Scott Zuyderduyn[7], Gary D. Bader[7], Scott VandenBerg[8], Lourdes Adriana Esparza[3], Marina Ryzhova[9], Wei Wang[6], Andrea Wittmann[1], Sebastian Stark[1], Laura Sieber[1], Huriye Seker-Cin[1], Linda Linke[1], Fabian Kratochwil[1], Natalie Jäger[10], Ivo Buchhalter[10], Charles D. Imbusch[11], Gideon Zipprich[11], Benjamin Raeder[4], Sabine Schmidt[12], Nicolle Diessl[12], Stephan Wolf[12], Stefan Wiemann[12], Benedikt Brors[10], Chris Lawerenz[11], Jürgen Eils[11], Hans-Jörg Warnatz[13], Thomas Risch[13], Marie-Laure Yaspo[13], Ursula D. Weber[6], Cynthia C. Bartholomae[14], Christof von Kalle[14,15], Eszter Turányi[16], Peter Hauser[17], Emma Sanden[18,19], Anna Darabi[18,19], Peter Siesjö[18,19], Jaroslav Sterba[20], Karel Zitterbart[20], David Sumerauer[21], Peter van Sluis[22], Rogier Versteeg[22], Richard Volckmann[22], Jan Koster[22], Martin U. Schuhmann[23], Martin Ebinger[23], H. Leighton Grimes[24], Giles W. Robinson[25,26], Amar Gajjar[26], Martin Mynarek[27], Katja von Hoff[27], Stefan Rutkowski[27], Torsten Pietsch[28], Wolfram Scheurlen[29], Jörg Felsberg[30], Guido Reifenberger[30], Andreas E. Kulozik[31], Andreas von Deimling[32], Olaf Witt[31], Roland Eils[10,15], Richard J. Gilbertson[25,26], Andrey Korshunov[32], Michael D. Taylor[5,33], Peter Lichter[6,15], Jan O. Korbel[4,34], Robert J. Wechsler-Reya[3] & Stefan M. Pfister[1,31]

Medulloblastoma is a highly malignant paediatric brain tumour currently treated with a combination of surgery, radiation and chemotherapy, posing a considerable burden of toxicity to the developing child. Genomics has illuminated the extensive intertumoral heterogeneity of medulloblastoma, identifying four distinct molecular subgroups. Group 3 and group 4 subgroup medulloblastomas account for most paediatric cases; yet, oncogenic drivers for these subtypes remain largely unidentified. Here we describe a series of prevalent, highly disparate genomic structural variants, restricted to groups 3 and 4, resulting in specific and mutually exclusive activation of the growth factor independent 1 family proto-oncogenes, *GFI1* and *GFI1B*. Somatic structural variants juxtapose *GFI1* or *GFI1B* coding sequences proximal to active enhancer elements, including super-enhancers, instigating oncogenic activity. Our results, supported by evidence from mouse models, identify *GFI1* and *GFI1B* as prominent medulloblastoma oncogenes and implicate 'enhancer hijacking' as an efficient mechanism driving oncogene activation in a childhood cancer.

Recent genome sequencing studies of medulloblastoma, a leading cause of cancer-related mortality in children[1], have yielded considerable insight into the genes, pathways and overall mutational landscape contributing to its pathogenesis[2–4]. Despite these advances, medulloblastoma remains a vastly heterogeneous disease characterized by very few recurrently mutated genes[5]. Medulloblastoma comprises at least four distinct molecular subgroups—wingless (WNT), sonic hedgehog (SHH), group 3 and group 4—each of which exhibits unique clinical and biological attributes, consistent with the concept of medulloblastoma existing not as a single entity, but more aptly a collection of different diseases[6,7].

Of the consensus subgroups, group 3 and 4 medulloblastomas have the poorest outcomes and remain least understood in terms of underlying genetics and biology[5]. Somatic *MYC* and *MYCN* amplifications rank among the most prevalent driver events known in these subgroups,

altered in just 17% (*MYC*) and 6% (*MYCN*) of group 3 and group 4 medulloblastomas, respectively[8]. Recurrent, somatically mutated genes are equally scarce, and for most cases, no obvious somatic 'drivers' have yet been revealed[5].

By analysing medulloblastoma genome sequencing data from different initiatives[2,4], we identified a series of spatially clustered somatic genomic structural variants (SVs) involving diverse SV classes that are linked to activation of *GFI1B* or its paralogue *GFI1* in group 3 and group 4 medulloblastomas. Further genomic and epigenomic analyses revealed a varied yet consistent interplay between SVs and the underlying epigenome that can explain *GFI1* and *GFI1B* activation in most cases. Functional analyses performed in mice confirmed the oncogenicity of GFI1 and GFI1B in the context of medulloblastoma. Collectively, these studies establish *GFI1* and *GFI1B* as novel, highly prevalent medulloblastoma oncogenes specifically activated in group 3 and group 4.

Diverse SVs activate GFI1B in medulloblastoma

Whole-genome sequencing (WGS; standard 100-base-pair (bp), paired-end and large-insert paired-end sequencing, see Methods) of 137 primary group 3 and 4 medulloblastoma samples (46 published[2,4] and 91 newly generated; Supplementary Table 1) facilitated a systematic, high-resolution screen for somatic SVs targeting novel medulloblastoma drivers. Rather than limiting our search to minimal common regions of recurrent amplification or deletion, a well-established approach for identifying somatically altered cancer genes[9,10], we considered all chromosomal rearrangements (that is, breakpoint clusters) detectable by WGS, including deletions, insertions, tandem duplications, amplifications, inversions and complex variants involving different SV classes (see Methods). Loci harbouring known medulloblastoma-related genes, including *MYCN* (2p24.3), *MYC* (8q24.21) and *SNCAIP* (5q23.2)[8], were readily recovered using this strategy (Fig. 1a). A novel prominent region of
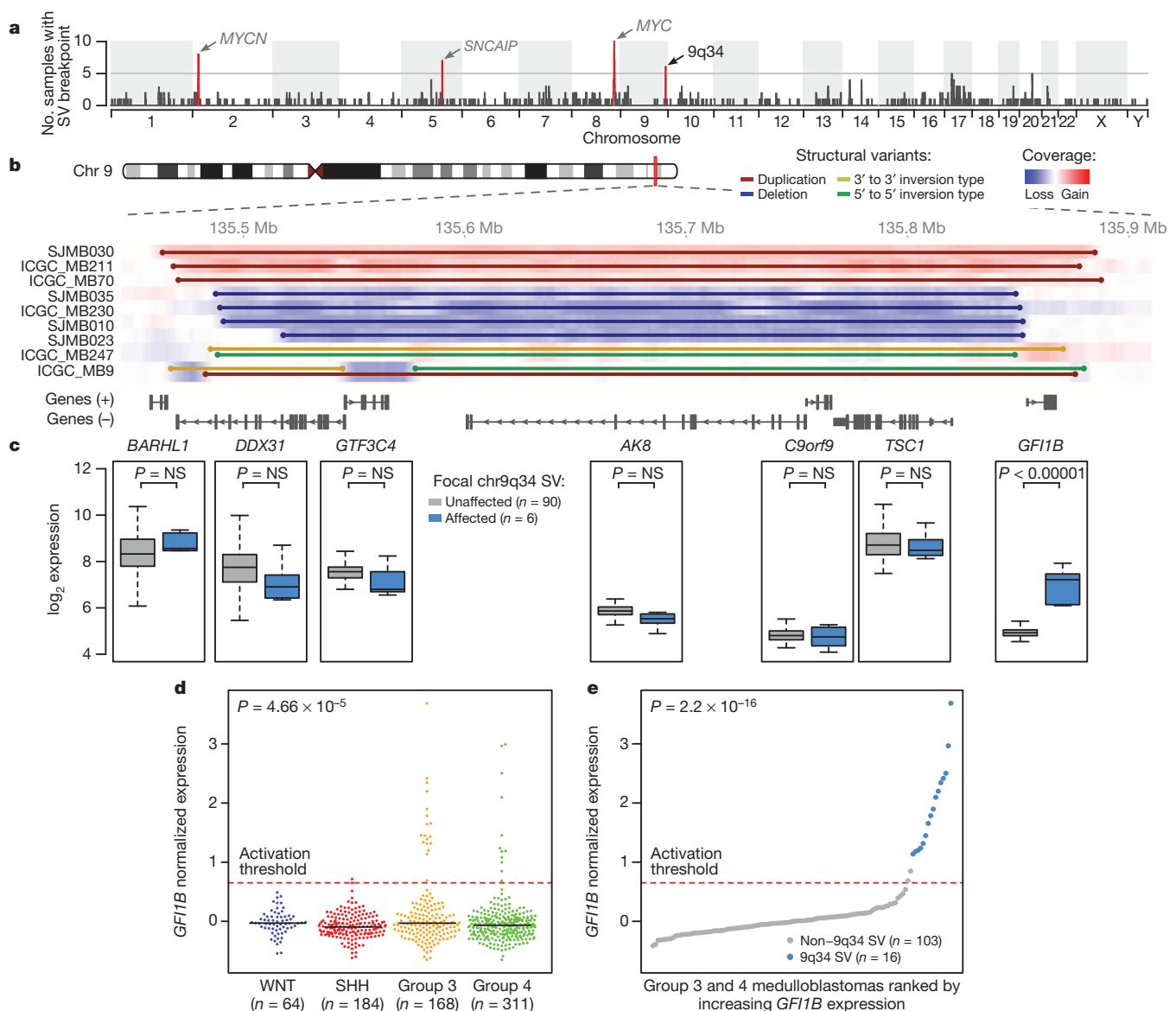
Figure 1 | Recurrent SVs activate the *GFI1B* proto-oncogene in medulloblastoma. **a**, Genome-wide SVs identified by WGS in a discovery cohort of group 3 and group 4 medulloblastomas (*n* = 137). **b**, Summary of SVs affecting a common locus of aberration on 9q34. **c**, Expression box-plots (*n* = 96) for the 7 genes contained within the 9q34 region of interest. Middle bar, median; lower and upper box limits, 25th and 75th percentiles, respectively; whiskers, 1.5 times the interquartile range from the 25th and 75th percentiles. **d**, *GFI1B* expression across medulloblastoma subgroups (*n* = 727). Dashed line delineates the threshold for detectable expression (see Methods). **e**, *GFI1B* expression for group 3 and group 4 medulloblastomas (*n* = 119) coloured according to 9q34 SV state. Dashed line indicates the threshold for detectable expression (see Methods).

interest mapped to chromosome 9q34.13 (Fig. 1a). Further assessment of our entire discovery series identified 9 of 137 (6.6%) cases with evidence of focal SV spanning this region of interest on chromosome 9 (135.46–135.89 megabases (Mb), ~425 kilobases (kb); Fig. 1b).

Instead of showing a predilection for a particular SV type, we observed a range of different SV classes at 9q34, including focal deletion (n = 4), tandem duplication (n = 3), and complex variants exhibiting inversions and focal deletions (n = 2). Examination of microarray-based copy-number data from our recent medulloblastoma genomics study[8] revealed additional evidence for subgroup-specific incidence of recurrent SVs affecting this region (Extended Data Fig. 1).

The region of interest on 9q34 harbours seven known genes (Fig. 1b), including the *TSC1* tumour suppressor gene previously implicated in medulloblastoma[11]. Integration of SV status with sample-matched gene expression data, however, uncovered highly specific transcriptional upregulation of *GFI1B* in samples harbouring 9q34 SV compared to non-affected counterparts (P < 0.00001, Fig. 1c). In contrast, neither *TSC1* nor any of the other remaining candidate genes exhibited a significant difference in expression in this context (Fig. 1c). Analysis of *GFI1B* expression in a large series of medulloblastomas (n = 727)[4,8] further substantiated this candidate, confirming restriction of *GFI1B* activation to groups 3 and 4, affecting 10.7% and 3.5% of cases from these subgroups, respectively (Fig. 1d).

To further characterize the relationship between somatic SVs at 9q34 and *GFI1B* transcriptional activation, we sequenced a validation set of 11 group 3 and group 4 medulloblastomas exhibiting *GFI1B* expression, confirming the existence of somatic SVs in 10 of 11 cases (Supplementary Table 2). In just one case (MAGIC_MB179), we failed to detect an underlying SV, suggesting that *GFI1B* overexpression might, in rare instances, be driven by an alternative non-SV-associated mechanism. Collectively, among 119 group 3 and group 4 medulloblastomas for which both WGS and matched expression array data were available, 16 of 18 (89%) *GFI1B*-activated cases displayed a detectable underlying SV (Fig. 1e).
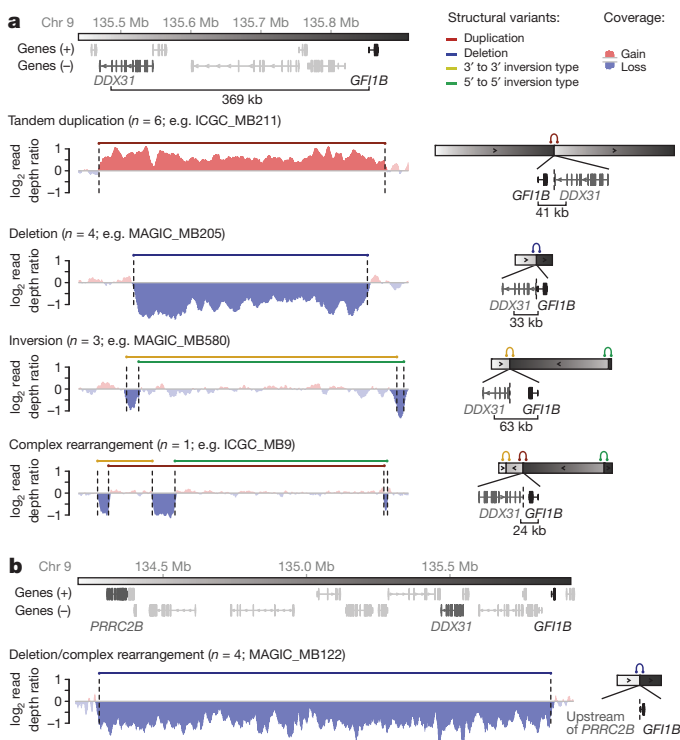
Importantly, every case showing SV at the 9q34 locus was associated with accompanying activation of *GFI1B* expression.

We next investigated each of the somatic SVs occurring at this locus in further detail to determine the mechanisms of *GFI1B* activation. Interestingly, irrespective of the underlying SV class, in 14 of 18 unique cases these events repositioned *GFI1B* proximal to the terminal sequence of the *DDX31* gene, a region that is normally positioned ~370 kb upstream of the *GFI1B* transcriptional start site (TSS) (Fig. 2a and Supplementary Table 2). Most affected samples annotated in our series juxtapose *GFI1B* within ~40 kb of *DDX31*, with the most distal introns of *DDX31* positioned either upstream or downstream of *GFI1B*, depending on the individual SV. Additionally, a smaller subset of examined cases (4 of 18) was found to exhibit broader deletions (~1.6 Mb) and complex rearrangements spanning a consistent region that starts upstream of the *PRRC2B* gene (chromosome 9, 134.27–134.28 Mb) and extends into the first intron (upstream of the first coding exon) of *GFI1B* (Fig. 2b).

The pattern of observed SVs does not support fusions of the *DDX31* coding sequence or its promoter with *GFI1B* as a common means of gene activation (see Fig. 2a). *DDX31–GFI1B* fusion transcripts were detected in two cases (2 of 4 *GFI1B*-activated medulloblastomas with available RNA-seq data: ICGC_MB9 and ICGC_MB247) but these were predicted to constitute non-functional (antisense or out-of-frame) alternative transcripts, not resulting in *GFI1B* activation (Extended Data Fig. 2).

Active enhancers drive GFI1B expression

The unexpected yet consistent observation of SVs resulting in juxtaposition of *GFI1B* to DNA elements normally located several hundred kilobases upstream suggested that rearrangements of *cis*-acting regulatory elements (such as enhancers) within these regions might be responsible for *GFI1B* activation. *DDX31* is highly expressed in group 3 and group 4 medulloblastomas and shows a correlated expression pattern with its two closest neighbours, *BARHL1* (downstream) and *GTF3C4* (upstream), suggesting that this locus exists in a transcriptionally permissive chromatin state in these medulloblastoma subgroups (Extended Data Fig. 3). To examine this locus and the surrounding region for evidence of enhancer activity, we used chromatin immunoprecipitation coupled with sequencing (ChIP-seq) for H3K27ac and H3K9ac, both known to mark active enhancers[12], in six primary group 3 medulloblastomas (Supplementary Table 1), including three *GFI1B*-activated cases (MAGIC_MB399, MAGIC_MB360 and ICGC_MB9; marked with an asterisk; Fig. 3a). Peak identification of these histone modification data predicted the presence of multiple enhancer clusters in this region, with peak H3K27ac and H3K9ac signals prominently overlapping or found immediately adjacent to the SV breakpoints observed in *GFI1B*-activated cases (Fig. 3a). Such clustering of highly active enhancers, and the overall H3K27 acetylation signal measured for these regions, is consistent with the recently described super-enhancers—regulatory elements associated with the expression of cell identity genes and master transcriptional regulators[13]. Super-enhancer identification (see Methods) performed on our H3K27ac ChIP-seq data suggested the presence of two such elements within the 9q34 region of interest (designated *PRRC2B* super-enhancer and *BARHL1/DDX31* super-enhancer, Fig. 3a and Extended Data Fig. 3).

GFI1B-activated medulloblastomas with SV breakpoints overlapping the inferred *BARHL1/DDX31* super-enhancer (MAGIC_MB360 and ICGC_MB9) showed markedly higher levels of H3K27ac and H3K9ac within this region (compared to non-*GFI1B*-activated samples), indicative of a potential feedback mechanism that increases the local enhancer signal (Fig. 3a). Moreover, H3K27ac and H3K9ac both mark the *GFI1B* locus in these two cases, suggesting 'spreading' of the activating enhancer marks from within the predicted super-enhancer to *GFI1B* consequent to genomic rearrangement (Fig. 3a). Allelic analysis of RNA-seq and enhancer ChIP-seq data for ICGC_MB9 demonstrated that both *GFI1B* expression and the active enhancer signals spanning *GFI1B* originate from the same allele (Fig. 3b), presumably the allele residing on the rearranged haplotype. Whole-genome bisulphite sequencing (WGBS)

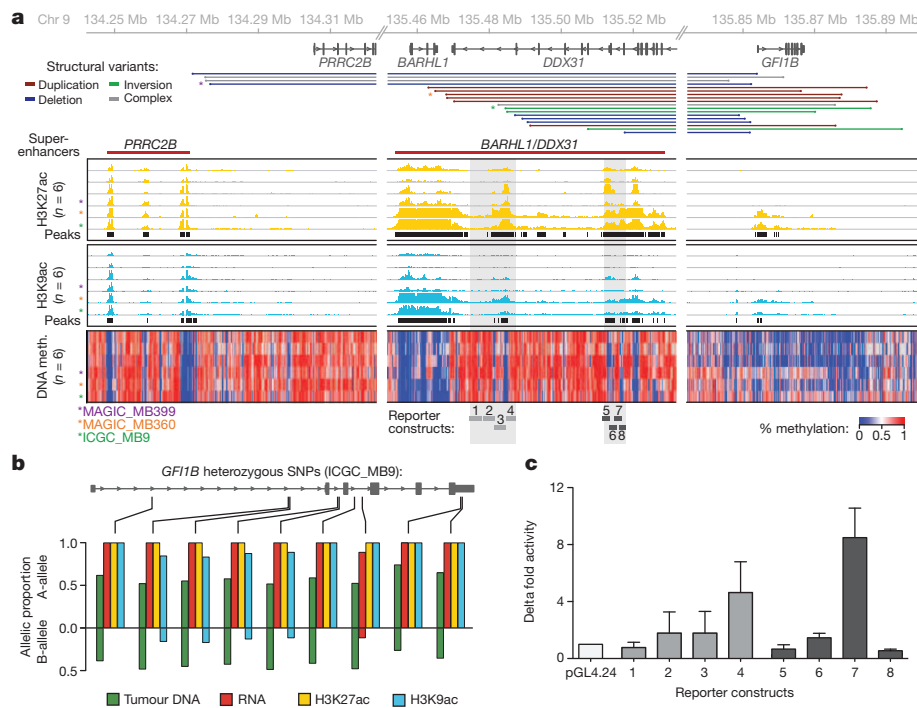
Figure 2 | Summary of recurrent SVs identified in *GFI1B*-activated medulloblastomas. a, b, Representative WGS coverage plots and associated schematics summarizing the different mechanisms of SV observed in *GFI1B*-activated medulloblastomas.

Figure 3 | Recurrent SVs juxtapose *GFI1B* proximal to active enhancers on 9q34. a, SV breakpoints (*n* = 18), enhancer-histone marks (H3K27ac and H3K9ac; *n* = 6) and whole-genome DNA methylation data (*n* = 6) overlapping the 9q34 locus in a subset of analysed medulloblastomas. **b**, Allelic analysis of RNA-seq and enhancer ChIP-seq reads overlapping *GFI1B*. **c**, Luciferase reporter activity for regions encompassed within the predicted enhancers indicated in panel **a** compared to empty vector. Error bars represent standard deviation from 3–4 independent experiments.

analysis of the same cases revealed profound DNA hypomethylation overlapping the putative enhancers identified by ChIP-seq, further supporting the accessibility of this chromatin to the transcriptional machinery in group 3 medulloblastomas (Fig. 3a).

To assess directly the capacity of identified enhancer elements to potentiate gene expression, a series of genomic fragments (~2 kb each) tiling two of the constituent enhancers (Fig. 3a, shaded regions) that contribute to the *BARHL1/DDX31* super-enhancer were independently tested for enhancer activity. Assays performed in the D425 group 3 medulloblastoma cell line confirmed robust reporter activity for constructs derived from either region, whereas constructs mapping outside of these peak regions failed to yield any detectable signal (Fig. 3c).

Mutually exclusive GFI1 and GFI1B activation

GFI1B is a paralogue of growth factor independence 1 (*GFI1*), with both genes functioning as SNAG-domain-containing zinc-finger transcriptional repressors essential for a variety of developmental processes, most notably haematopoiesis[14–16]. Importantly, extensive mouse genetics and insertional mutagenesis screens have established *Gfi1* and *Gfi1b* as potent proto-oncogenes in subtypes of leukaemias and lymphomas[17,18]. However, no recurrent somatic SVs affecting *GFI1* or *GFI1B* have been reported in these or any other cancers. Transcriptional analysis showed clear activation of *GFI1* in a subset of medulloblastomas (29 of 724, 4.0%), with expression tightly restricted to group 3 medulloblastomas ($P < 2 \times 10^{-16}$; Fig. 4a). Comparison of *GFI1* and *GFI1B* expression
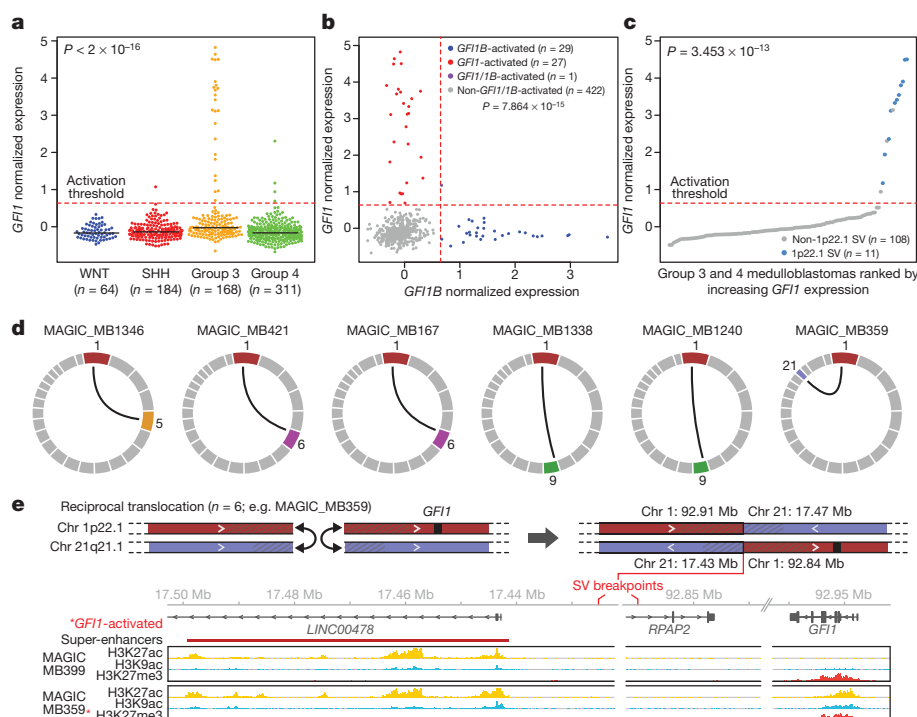
Figure 4 | Mutually exclusive activation of *GFI1* and *GFI1B* in medulloblastoma. a, b, *GFI1* expression is largely restricted to group 3 (**a**) and is mutually exclusive from *GFI1B* expression (**b**). **c**, *GFI1* expression for group 3 and group 4 medulloblastomas (*n* = 119) coloured according to underlying SV state. Dashed line indicates the threshold for detectable gene expression (see Methods). **d**, Summary of *GFI1* translocations (*n* = 6) observed in group 3 medulloblastoma. **e**, Schematic of the reciprocal t(1:21) translocation observed in *GFI1*-activated MAGIC_MB359. Histone marks overlapping the breakpoints proximal to *GFI1* and the partner chromosome 21 translocation region are shown for a non-*GFI1*-activated case (MAGIC_MB399) and the translocation case (MAGIC_MB359).

among group 3 and group 4 medulloblastomas showed a mutually exclusive pattern of activation ($P = 7.864 \times 10^{-15}$; Fig. 4b), further supportive of their oncogenic roles in medulloblastoma. Collectively, *GFI1* and *GFI1B* expression was observed in 25% and 5% of group 3 and group 4 medulloblastomas, respectively (Extended Data Fig. 4). These findings were validated in an independent series of medulloblastomas ($n = 156$) by immunohistochemical analysis, confirming mutually exclusive GFI1 and GFI1B expression that contributed to 41% and 10% of group 3 and group 4 cases analysed, respectively (Extended Data Fig. 4).

GFI1/GFI1B-expressing medulloblastomas did not form their own discrete subtype within the group 3 subgroup as evaluated by clustering of either gene expression or DNA methylation data (Extended Data Fig. 5). Activation status of GFI1 and GFI1B was associated with patient age in group 3, occurring exclusively in non-infant cases in the gene expression cohort ($P < 0.0001$, chi-squared test; Extended Data Fig. 5). However, no association with patient outcome or other clinical/demographic variables was observed in either our combined gene expression or formalin-fixed paraffin-embedded cohorts (Extended Data Fig. 5).

To investigate whether *GFI1* activation in medulloblastoma is attributable to SV mechanisms similar to those targeting *GFI1B*, we examined the *GFI1* locus in our discovery WGS series of group 3 and group 4 medulloblastomas ($n = 137$) and sequenced an additional validation set consisting of 11 non-overlapping *GFI1*-activated cases. This strategy revealed a diversity of SVs affecting the *GFI1* locus or surrounding genomic regions, including interchromosomal translocations ($n = 6$), tandem duplications ($n = 4$) and a complex rearrangement ($n = 1$), respectively, in medulloblastomas exhibiting *GFI1* expression (Fig. 4d, e and Extended Data Fig. 6). We confirmed somatic SVs in 11 of 14 *GFI1*-activated cases analysed (Fig. 4c), demonstrating that, similar to *GFI1B*, *GFI1* activation is typically associated with an underlying SV.

RNA-seq analysis did not disclose evidence for possible *GFI1* fusion genes (data not shown), suggesting that the detected rearrangements contribute to *GFI1* activation by alternative mechanisms. Observed translocation partners showed no apparent preference for intragenic or intergenic breakpoints (Supplementary Table 2). Overlaying histone ChIP-seq data with translocation breakpoint regions revealed activating enhancer-histone modification states close to the observed breakpoints (Fig. 4e and Extended Data Fig. 7), suggesting translocations of the normally repressed *GFI1* locus into actively transcribed regions as

the likely mechanism of gene activation. Importantly, most *GFI1* translocation partners were confirmed to harbour clusters of highly active enhancers consistent with super-enhancers that were situated proximal to sequenced breakpoints, analogous to what we observed for *GFI1B*-activated cases (Fig. 4e and Extended Data Fig. 7).

Despite identifying multiple distinct t(1:6) and t(1:9) translocations (Fig. 4d), the only recurrent SV detected in *GFI1*-activated medulloblastomas was a focal (~6 kb) tandem duplication located ~45 kb downstream of *GFI1*, identified in three *GFI1*-activated samples but not in any other sequenced sample (Extended Data Figs 6 and 7). Enhancer mark ChIP-seq analysis confirmed that this focal region was profoundly marked by the active H3K27ac mark in the context of tandem duplication but not in non-activated group 3 medulloblastomas (Extended Data Fig. 7), suggesting that this region downstream of *GFI1* can, when duplicated, promote its activation.

GFI1 and GFI1B promote medulloblastoma formation *in vivo*

Mouse models of medulloblastoma have given important insights into disease biology[19,20]. Recently, two group 3 models have been described[21,22]. Each of these involves overexpression of *Myc* with *Trp53* loss-of-function—a combination not typically observed in human medulloblastomas, as *MYC* amplification/overexpression (group 3) and *TP53* mutations (WNT and SHH subgroups) occur in different subgroups[5,23].

Group 3 medulloblastoma expression data confirmed significant upregulation of *MYC* in *GFI1*-activated cases versus non-*GFI1/GFI1B*-activated, subgroup-matched counterparts (Extended Data Fig. 8). Pathway analysis identified MYC target gene sets as being highly enriched in *GFI1/GFI1B*-activated group 3 medulloblastomas (Extended Data Fig. 8). Additionally, co-occurrence of *MYC* amplification and *GFI1* activation was noted in a subset of group 3 medulloblastomas (Extended Data Fig. 8), further suggesting that GFI1 and MYC may cooperate to promote group 3 medulloblastoma. Indeed, *Gfi1* and *Myc* are known to function as synergistic oncogenes and enhance T-cell lymphomagenesis in transgenic mouse models[24,25].

To evaluate whether *GFI1* and *GFI1B* can function as novel medulloblastoma oncogenes, we used an orthotopic transplantation model[21] whereby retroviruses encoding GFI1 or GFI1B were transduced either alone or in combination with viruses encoding MYC into neural stem cells, followed by their transplantation into the cerebella of immunocompromised mice (Fig. 5a). Neither GFI1 nor GFI1B alone was sufficient to promote
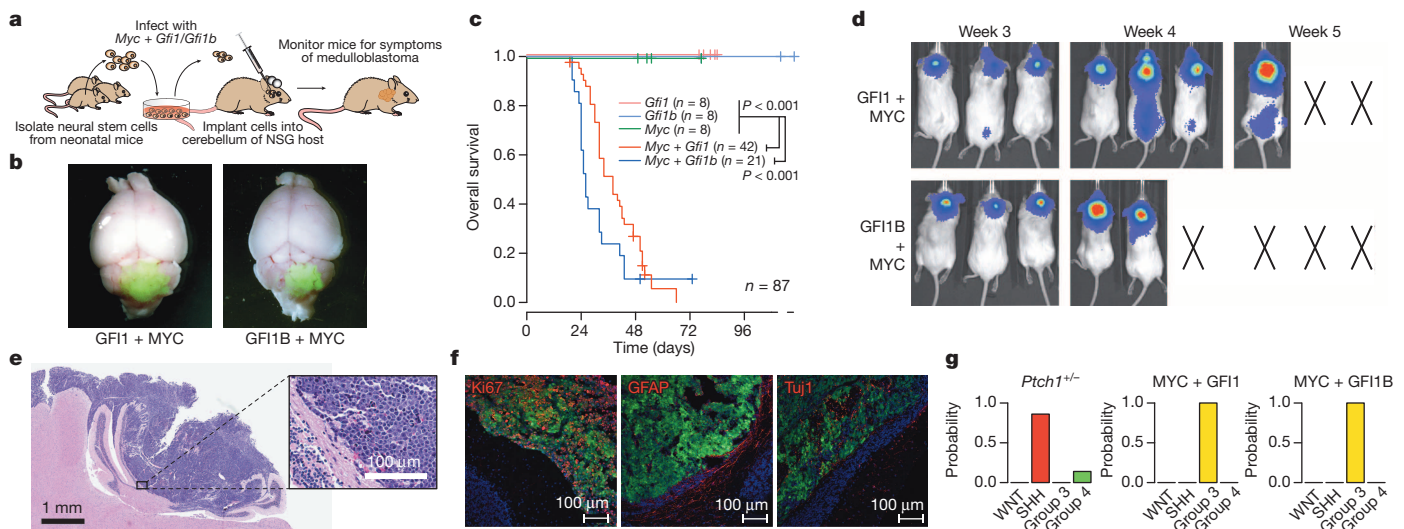
Figure 5 | GFI1 and GFI1B cooperate with MYC to promote medulloblastoma formation in mice. a, Strategy for evaluating *Gfi1* and *Gfi1b* as putative medulloblastoma oncogenes. **b**, Whole-mount images of GFP-expressing MYC + GFI1 (left) and MYC + GFI1B (right) tumours. **c**, Survival curves for animals receiving 1×10^5 cells infected with viruses carrying the indicated transgenes. **d**, Bioluminescent imaging of recipient animals at the

indicated time points. X symbols denote animals that were euthanized before reaching the indicated time point. **e**, Haematoxylin and eosin staining of cerebellar sections derived from MYC + GFI1 tumour-bearing mice. **f**, Immunofluorescence imaging of MYC + GFI1 tumours stained with the indicated antibodies. **g**, Subgroup probabilities for *Ptch1*$^{+/-}$, MYC + GFI1, and MYC + GFI1B models based on cross-species molecular classification.
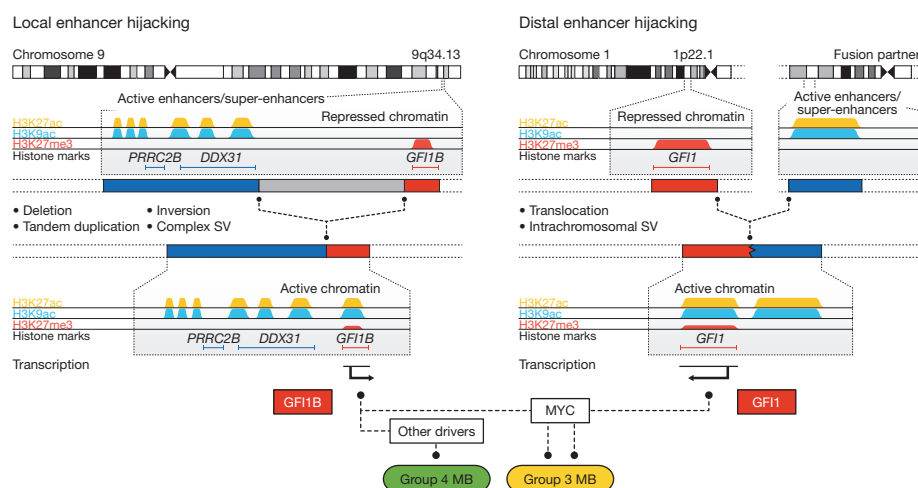
Figure 6 | Summary of inferred mechanisms underlying *GFI1* and *GFI1B* activation in medulloblastoma. Predominant mechanisms of SV and corresponding genomic redistribution of active enhancers, including super-enhancers, observed in *GFI1/GFI1B*-activated medulloblastomas. Activation of *GFI1* and *GFI1B* occurs in a mutually exclusive manner in either group 3 or group 4 and both oncogenes can cooperate with MYC to promote medulloblastoma pathogenesis.

tumorigenesis in this system (Fig. 5c and Extended Data Fig. 9). When combined with MYC, however (which is also insufficient to generate medulloblastoma on its own in this system[21]), both GFI1 (that is, MYC + GFI1) and GFI1B (that is, MYC + GFI1B) rapidly produced highly aggressive cerebellar tumours in nearly all recipient mice within 4–5 weeks ($n = 37$ of 42 and 19 of 21 with median survival time of 38 days and 26 days for MYC + GFI1 and MYC + GFI1B, respectively; Fig. 5b–f).

Cerebellar sections derived from either MYC + GFI1 or MYC + GFI1B recipient mice showed large masses of infiltrating tumour cells with marked cellular pleomorphism, morphologically consistent with large cell, anaplastic (LCA) medulloblastoma (Fig. 5e and Extended Data Fig. 9). LCA histology is significantly more prevalent in group 3 medulloblastoma (∼20–25% of cases) than in other medulloblastoma subgroups[6,7]. Metastatic dissemination was also noted in 30–50% of MYC + GFI1 and MYC + GFI1B tumour-bearing mice (Fig. 5d and data not shown), paralleling the high frequency of metastasis seen in group 3 medulloblastoma patients[6,7]. Moreover, immunofluorescence microscopy confirmed that MYC + GFI1 and MYC + GFI1B tumours are highly proliferative and express neuronal but not glial lineage markers (Fig. 5f and Extended Data Fig. 9), consistent with a medulloblastoma-like immunophenotype. Transcriptional profiling and subsequent multidimensional scaling analysis demonstrated a notable similarity between the GFI1- and GFI1B-driven models and confirmed an expression signature consistent with human group 3 medulloblastoma counterparts, suggesting that these models recapitulate molecular characteristics of the human disease (Fig. 5g and Extended Data Fig. 9).

Discussion

Medulloblastoma sequencing studies have highlighted the intertumoral molecular heterogeneity underpinning this malignancy, revealing very few recurrently mutated driver genes, especially in group 3 and group 4 (ref. 5). Here, we have identified somatic genomic rearrangements in association with mutually exclusive *GFI1* and *GFI1B* activation in approximately one-third of group 3 medulloblastomas—now qualifying these oncogenes as the most prevalent drivers in this subgroup (Fig. 6). Moreover, 5–10% of group 4 medulloblastomas harbour analogous SV associated with *GFI1/GFI1B* activation, reinforcing the notion that these subgroups share some biological similarities[6,7].

The verification of diverse SVs in nearly all *GFI1/GFI1B*-activated medulloblastomas analysed in this study has implications for future cancer genome studies. Conventional approaches for identifying genes recurrently targeted by SV in cancer usually focus on minimal common regions of aberration and require that putative gene targets are (at least partially) included within these altered regions[9]. In contrast to the high-level amplifications known to target *MYC*, *MYCN* and other recognized medulloblastoma oncogenes[3,8], *GFI1* and *GFI1B* are not amplified in medulloblastoma. Observations extracted from the current study revealed that (1) a considerable proportion of SVs leading to *GFI1* and *GFI1B* activation do not actually include the target gene itself; and (2) multiple distinct classes of SV including duplication, deletion, inversion and other complex rearrangements can converge on activation of a single target, often without associated gene-level copy-number change. Our findings suggest that similar mechanisms leading to gene deregulation (that is, activation of oncogenic drivers) might have thus far been overlooked in other cancers.

SV-dependent redistribution of *GFI1* and *GFI1B* from regions of transcriptionally silent chromatin to regions populated with active enhancers, such as super-enhancers (Fig. 6), underscores the diversity of interplay between the cancer genome and epigenome[26–28]. Activation of *GFI1* and *GFI1B* seemingly does not rely on specific epigenetic deregulation but rather implicates a form of 'enhancer hijacking' whereby oncogene activation hinges on the appropriation of a physiologically active epigenetic state from proximal or distant loci, including those mapping to other chromosomes. This concept of merging oncogenes with active regulatory elements has long been observed in lymphoid malignancies, where translocations are known to relocate *MYC*, *BCL2* and other oncogenes adjacent to highly active promoter or enhancer loci, most commonly those belonging to the immunoglobulin genes (that is, IgH/IgL loci) or T-cell receptors (that is, TCR-α/β loci)[29]. To the best of our knowledge, this is the first report to substantiate such a phenomenon in brain tumours.

We have discovered a series of highly variable genomic rearrangements leading to oncogene activation in a significant proportion of cases from poorly understood medulloblastoma subgroups, implicating *GFI1* and *GFI1B* as novel oncogenic drivers worthy of pursuit as candidates for molecularly targeted therapy. The patterns of rearrangement associated with *GFI1* and *GFI1B* activation described here have broad-reaching implications for cancer genomics, and warrant the implementation of similar efforts to revisit existing sequencing data using analytical approaches that extend beyond the coding genome. On the basis of our observations, it is tempting to speculate that similar 'enhancer hijacking' may be equally prevalent in other solid cancers.

METHODS SUMMARY

All patient material included in this study was collected after receiving informed consent from the patients and their families. Medulloblastoma samples were collected at first resection, before adjuvant chemotherapy or radiotherapy. Full details on the sequencing cohorts included in this report are summarized in Supplementary Table 1. Medulloblastoma subgroups were assigned using gene expression array data, a custom nanoString CodeSet, DNA methylation profiling, or a combination of the above, as previously described[7,30,31]. WGS and long-range paired-end mapping were performed as described[2,3]. WGBS and DNA methylation analysis was conducted as described[32]. Chromatin extraction, immunoprecipitation and library preparation for ChIP-seq studies was performed using proprietary methods at Active Motif (Carlsbad). H3K27ac and H3K9ac peaks were called using

BayesPeak[33]. Super-enhancers were inferred using the ROSE algorithm with default parameters as described[28]. Affymetrix expression array profiling of human and mouse tumour RNAs was performed at core facilities within the Amsterdam Medical Centre (Amsterdam, Netherlands), German Cancer Research Center (Heidelberg, Germany), and The Hospital for Sick Children (Toronto, Canada). Mouse studies were conducted at the Sanford-Burnham Medical Research Institute and Sanford Consortium for Regenerative Medicine Animal Facilities in accordance with national regulations using procedures approved by the Institutional Animal Care and Use Committees at Sanford-Burnham and the University of California San Diego.

1. Ostrom, Q. T. et al. CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2006–2010. Neuro-oncol. 15 (Suppl 2), 1–56 (2013).
2. Jones, D. T. et al. Dissecting the genomic complexity underlying medulloblastoma. Nature 488, 100–105 (2012).
3. Rausch, T. et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell 148, 59–71 (2012).
4. Robinson, G. et al. Novel mutations target distinct subgroups of medulloblastoma. Nature 488, 43–48 (2012).
5. Northcott, P. A. et al. Medulloblastomics: the end of the beginning. Nature Rev. Cancer 12, 818–834 (2012).
6. Cho, Y. J. et al. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. J. Clin. Oncol. 29, 1424–1430 (2011).
7. Northcott, P. A. et al. Medulloblastoma comprises four distinct molecular variants. J. Clin. Oncol. 29, 1408–1414 (2011).
8. Northcott, P. A. et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. Nature 488, 49–56 (2012).
9. Santarius, T., Shipley, J., Brewer, D., Stratton, M. R. & Cooper, C. S. A census of amplified and overexpressed human cancer genes. Nature Rev. Cancer 10, 59–64 (2010).
10. Kim, T. M. et al. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. Genome Res. 23, 217–227 (2013).
11. Bhatia, B. et al. Tuberous sclerosis complex suppression in cerebellar development and medulloblastoma: separate regulation of mammalian target of rapamycin activity and p27 Kip1 localization. Cancer Res. 69, 7224–7234 (2009).
12. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473, 43–49 (2011).
13. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319 (2013).
14. Hock, H. et al. Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. Nature 431, 1002–1007 (2004).
15. Person, R. E. et al. Mutations in proto-oncogene GFI1 cause human neutropenia and target ELA2. Nature Genet. 34, 308–312 (2003).
16. Saleque, S., Cameron, S. & Orkin, S. H. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. Genes Dev. 16, 301–306 (2002).
17. Gilks, C. B., Bear, S. E., Grimes, H. L. & Tsichlis, P. N. Progression of interleukin-2 (IL-2)-dependent rat T cell lymphoma lines to IL-2-independent growth following activation of a gene (Gfi-1) encoding a novel zinc finger protein. Mol. Cell. Biol. 13, 1759–1768 (1993).
18. Scheijen, B., Jonkers, J., Acton, D. & Berns, A. Characterization of pal-1, a common proviral insertion site in murine leukemia virus-induced lymphomas of c-myc and Pim-1 transgenic mice. J. Virol. 71, 9–16 (1997).
19. Gibson, P. et al. Subtypes of medulloblastoma have distinct developmental origins. Nature 468, 1095–1099 (2010).
20. Goodrich, L. V., Milenkovic, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. Science 277, 1109–1113 (1997).
21. Pei, Y. et al. An animal model of MYC-driven medulloblastoma. Cancer Cell 21, 155–167 (2012).
22. Kawauchi, D. et al. A mouse model of the most aggressive subgroup of human medulloblastoma. Cancer Cell 21, 168–180 (2012).
23. Zhukova, N. et al. Subgroup-specific prognostic implications of TP53 mutation in medulloblastoma. J. Clin. Oncol. 31, 2927–2935 (2013).
24. Zornig, M., Schmidt, T., Karsunky, H., Grzeschiczek, A. & Moroy, T. Zinc finger protein GFI-1 cooperates with myc and pim-1 in T-cell lymphomagenesis by reducing the requirements for IL-2. Oncogene 12, 1789–1801 (1996).
25. Schmidt, T. et al. Zinc finger protein GFI-1 has low oncogenic potential but cooperates strongly with pim and myc genes in T-cell lymphomagenesis. Oncogene 17, 2661–2667 (1998).
26. Plass, C. et al. Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. Nature Rev. Genet. 14, 765–780 (2013).
27. Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. Cell 153, 38–55 (2013).
28. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. Cell 155, 934–947 (2013).
29. Nambiar, M., Kari, V. & Raghavan, S. C. Chromosomal translocations in cancer. Biochim. Biophys. Acta 1786, 139–152 (2008).
30. Hovestadt, V. et al. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. Acta Neuropathol. 125, 913–916 (2013).
31. Northcott, P. A. et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol. 123, 615–626 (2012).
32. Hovestadt, V. et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature http://dx.doi.org/10.1038/nature13268 (18 May 2014).
33. Cairns, J. et al. BayesPeak–an R package for analysing ChIP-seq data. Bioinformatics 27, 713–714 (2011).

Author Contributions P.A.N., C.L., T.Z., A.M.S., D.K., L.A.E., W.W., A.W., S.St., L.S., H.S.-C., L.L., F.K., J.F., B.R., S.Sc., N.D., S.Wo., T.R., C.C.B., P.v.S. and A.K. performed and/or coordinated experimental or technical work. P.A.N., T.Z., S.E., D.J.H.S., V.H., M.Z., S.Z., G.D.B., N.J., I.B., C.D.I., G.Z., J.E., R.Vo., J.K. and J.O.K. performed and/or coordinated data analysis. M.Re., F.M.G.C., S.V., M.Ry., E.T., P.H., E.S., A.D., P.S., J.S., K.Z., D.Su., M.U.S., M.E., H.L.G., G.W.R., A.G., M.M., K.v.H., S.R., T.P., W.S., R.J.G., A.K. and M.D.T. contributed data, provided reagents, or patient materials. P.A.N., C.L., T.Z., S.E., D.J.H.S., V.H., D.St., D.T.W.J., M.K., S.Z., H.-J.W., R.J.G., M.D.T., P.Li., J.O.K., R.J.W.-R. and S.M.P. prepared the initial manuscript and display items. P.A.N., G.D.B., S.Wi., B.B., C.L., M-L.Y., U.D.W., C.v.K., R.V., G.R., A.E.K., A.v.d., O.W., R.E., P.Li., J.O.K., R.J.W.-R. and S.M.P. provided project leadership. P.A.N., J.O.K., R.J.W.-R. and S.M.P. co-conceived and led the study. P.Li., J.O.K., R.J.W.-R. and S.M.P are co-senior authors of this study.

Author Information Short-read sequencing data have been deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/) hosted by the EBI, under accession number EGAS00001000215. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M.P. (s.pfister@dkfz-heidelberg.de), R.J.W.-R. (rwreya@sanfordburnham.org), J.O.K. (korbel@embl.de) or P.Li. (m.macleod@dkfz-heidelberg.de).

METHODS

General statistical methods. All statistical tests were performed in the R Statistical Environment (R version 3.0.0) unless otherwise specified. The Kolmogorov–Smirnov test was used to compare candidate gene expression in chr9q34 SV cases to non-SV cases. Differential expression of *GFI1* and *GFI1B* across medulloblastoma subgroups was calculated using ANOVA. Enrichment of underlying locus-specific SVs in *GFI1/GFI1B*-expressing cases was calculated using Fisher's exact test. Mutual exclusivity of *GFI1* and *GFI1B* expression in group 3 and group 4 medulloblastomas was determined using Fisher's exact test. Survival analyses were performed in GraphPad Prism 5 using the log-rank (Mantel-Cox) test to compare survival differences between groups.

Sample collection and preparation. An Institutional Review Board ethical vote (Ethics Committee of the Medical Faculty of Heidelberg) was obtained according to ICGC guidelines (http://www.icgc.org), along with informed consent for all participants. No patient underwent chemotherapy or radiotherapy before surgical removal of the primary tumour. Tumour tissues were subjected to neuropathological review for confirmation of histology and for tumour cell content >80%. Analytes were isolated as previously described[2]. Cells were cultured at 37 °C with 5% CO_2. D425_Med MB cells (D425; a gift from Professor Darrell D. Bigner) were cultured in DMEM with 10% FCS (Life Technologies) and regularly authenticated and tested for mycoplasma (Multiplexion, Heidelberg, Germany). Validation samples for WGS were obtained in accordance with the Research Ethics Board at The Hospital for Sick Children (Toronto, Canada).

High-throughput sequencing data generation. *Short-insert paired-end sequencing.* Samples were processed and libraries sequenced as previously described[2].

Medulloblastoma and germline WGS data[4] generated by the Paediatric Cancer Genome Project (http://explore.pediatriccancergenomeproject.org/) was accessed from The European Genome-phenome Archive (Study ID EGAS00001000347). The original alignments of this WGS data were performed against either reference genome hg18 or hg19. For comparability with our data, the alignment files in hg18 have been converted to FASTQ files using Picard tools (http://picard.sourceforge.net) providing the 'SamToFastq' option. For the alignment of the FASTQ files, the same reference genome as used in the creation of the original hg19 BAM files has been used along with BWA for alignment and Picard for merging and duplicate read filtering.

Long-range paired-end sequencing data generation. Long-range (or 'Mate-pair') DNA library preparation was carried out as previously described[2] or using the newer Nextera Mate Pair Sample Preparation Kit (Illumina). In brief, 4 µg of high-molecular-mass genomic DNA were fragmented by the Tagmentation reaction in 400 µl, followed by the strand displacement and AMPure XP (Agencourt) clean-up reaction. Samples were size selected to 4–6 kb with a gel step following the Gel-Plus path of the protocol. 300–550 ng of size-selected DNA were circularized in 400 µl for 16 h at 30 °C. The library was then constructed after an exonuclease digestion step to get rid of remaining linear DNA, fragmentation to 300–700 bp with a Covaris S2 instrument (LGC Genomics), binding to streptavidin beads and Illumina Truseq adaptor ligation. Final library was obtained after PCR for 1 min at 98 °C, followed by 9 cycles of 30 s at 98 °C, 30 s at 60 °C, 1 min at 72 °C and a final 5 min at 72 °C step. Deep sequencing was carried out with the Illumina HiSeq2000 (2 × 101 bp) instrument to reach an average physical coverage of 20–30×.

ChIP sequencing. Chromatin extraction, immunoprecipitation and library preparation for ChIP-seq were performed at Active Motif (Carlsbad, CA) according to proprietary methods. Briefly, 15 µg of chromatin were used as input for ChIP with ChIP-grade antibodies recognizing H3K27ac (AM#39133, Active Motif), H3K9ac (AM#39918, Active Motif), or H3K27me3 (#07-449, Millipore). Libraries were sequenced on the Illumina HiSeq 2000 platform using 2 × 101 cycles according to the manufacturer's instructions.

Whole-genome bisulphite sequencing. Whole-genome bisulphite library preparation was carried out as recently described[34], with modifications to a previously published protocol[35]. In brief, 5 µg of genomic DNA were sheared using a Covaris device (Covaris Inc.). After adaptor ligation, DNA fragments with insert lengths of 200–250 bp were isolated using an E-Gel electrophoresis system (Life Technologies) and bisulphite converted using the EZ DNA Methylation kit (Zymo Research). PCR amplification of the fragments was performed in six parallel reactions per sample using the FastStart High Fidelity PCR kit (Roche). Library aliquots were then pooled per sample and sequenced on an Illumina HiSeq 2000 machine.

RNA sequencing. RNA quality control was performed using the 2100 Bioanalyzer platform (Agilent). RNA sequencing libraries were prepared using the TruSeq stranded protocol with Ribo-Zero Gold (Illumina) and sequenced on the Illumina HiSeq 2000 platform with 2 × 51 cycles according to the manufacturer's instructions.

High-throughput sequencing data analysis. *Whole-genome sequencing.* Short-insert WGS data was analysed as previously described[2]. Long-range paired-end sequencing reads were aligned to the hg19 assembly of the human reference genome using the Illumina-provided alignment software (ELAND, version 2).

Structural variant discovery and filtering. Deletions, tandem duplications, inversion, translocations, as well as complex rearrangements resulting in the corresponding paired-end signatures were inferred using DELLY v0.0.11 (ref. 36). We considered all those predictions as somatic that were not present in a set of 1,000 Genomes Project (1000GP; http://1000genomes.org)[37] samples corresponding to germline samples taken from normal healthy individuals. Specifically, we used DELLY to infer variants in 1,106 samples belonging to phase 1 of the 1000GP. Furthermore, we inferred variants in the germline samples belonging to the studied tumours. For a given tumour sample, we considered all those variants as somatic that were present neither in any of the 1000GP samples nor in any of the additional germline samples. Two SVs were considered as identical if their start and end coordinates differed by less than 5.0 kb (approximate insert size of a long-range paired-end library) and if their reciprocal overlap was larger than 50%. Variants that were present in the control samples were either true germline variants or represented artefacts caused by misalignment of reads (for example, due to inaccuracies within the human reference genome). To consider a variant prediction as high-confidence we further required at least four supporting read pairs with a minimum median mapping quality of 20 for each event to exclude false-positive predictions caused by randomly mapping low-quality reads.

Region identification. We divided the human reference genome into overlapping 1-Mb windows (100-kb offset). For each window, we counted the number of samples with at least one SV breakpoint in the given region (based on short-insert as well as long-insert paired-end sequencing data). Only focal high-confidence SV predictions were used in this analysis (20 kb to 10 Mb in size). Windows affected in at least five samples were investigated manually.

Copy-number analysis. We determined the number of sequencing reads per non-overlapping genomic window of size 250 bp (high-coverage paired-end data) or 1,000 bp (low-coverage long-range paired-end data) for tumour samples with chr9q34 or chr1p22 SV and their corresponding controls. Tumour values were normalized by the ratio of read counts between tumour and controls within a 500-kb region. Subsequently, for each window, the $\log_2$ ratio between normalized tumour and control counts was determined. These values were averaged along a sliding window of 5 kb (short-insert paired-end data) or 10 kb (long-range paired-end data). For tumour samples without a matching control sample, the control of ICGC_MB230 was used.

ChIP sequencing. Histone ChIP-seq data for H3K27ac, H3K9ac and H3K27me3 was processed by the Illumina analysis pipeline (version 1.8.3) and aligned to the Human Reference Genome (assembly hg19, GRCh37) using BWA version 0.5.9-r16 (ref. 38). Putative PCR duplicates were filtered using Picard MarkDuplicates (http://picard.sourceforge.net). For downstream analyses, we generated whole-genome coverage tracks with reads normalized to all properly paired reads (RPM, paired-end reads/fragments per million). We used igvtools version 2.2.2 (http://www.broadinstitute.org/igv/igvtools) and the non-default parameter–pairs and a window size of 25. For peak-calling of histone marks, ChIP-seq data for each histone modification (H3K27ac or H3K9ac) was used to generate individual BED files for analysed samples using BEDTools[39]. Individual BED files were then combined for each histone modification and peaks were called using the Bioconductor BayesPeak package in R[33]. Super-enhancers were identified using the ROSE algorithm with default parameters (stitching distance of 12,500 bp and promoter exclusion region of ±2,000 bp around TSS)[28]. Briefly, peaks called via BayesPeak were used as constituent enhancers to run the algorithm and super-enhancers were called by ranking of H3K27ac signal at stitched constituent enhancers.

Whole-genome bisulphite sequencing. WGBS sequencing data was analysed using methylCtools (Hovestadt *et al.*, manuscript in preparation). In brief, methylCtools builds on BWA and adds functionality for aligning bisulphite treated DNA to a reference genome in a similar manner as described previously[40]. Sequencing reads were adaptor-trimmed using SeqPrep (https://github.com/jstjohn/SeqPrep) and translated to a fully C-to-T converted state. Alignments were performed against a single index of both *in silico* bisulphite-converted strands of the human reference genome (hg19, NCBI build 37.1) using BWA version 0.6.1-r104 (ref. 38) and the non-default parameters -q 20 -s. Previously translated bases were translated back to their original state, and reads mapping antisense to the respective reference strand were removed. Putative PCR duplicates were filtered using Picard MarkDuplicates (http://picard.sourceforge.net). Non-conversion rates were estimated on the basis of lambda phage genome spike-ins. Single base pair methylation ratios (beta-values) were determined by quantifying evidence for methylated (unconverted) and unmethylated (converted) cytosines at all CpG positions. Only properly paired or singleton reads with mapping quality of ≥1 and bases with Phred-scaled quality score of ≥20 were considered. To account for population variability, we filtered CpGs for which more than 25% of reads at a given position (on either strand) were not supportive of this CpG being in fact a CpG in the sample being analysed. Subsequently, information from both strands was combined and CpGs with coverage less than five reads were set as NA.

RNA sequencing. Demultiplexed FASTQ files were generated using the Bcl2FastQ conversion software (Illumina, version 1.8.4). The resulting sequencing reads were aligned to the human genome reference build hg19 (version human_g1k_v37 – 1,000 Genomes Project Phase 1) using BWA version 0.5.9-r16 (ref. 38) with default parameters. Only the chromosomes 1–22, X, Y and M were used for the mapping. Read coverage plots were prepared using the UCSC Genome Browser showing the number of aligned reads for each genomic position per million mapped reads (RPM) with mapping quality MAPQ >1. The sequencing reads were also used as input for the TopHat2-Fusion algorithm[41] for detection of gene fusion breakpoints.

Allelic analysis. Germline SNPs were determined using Samtools and BCFtools. For each SNP, the number of reads in the tumour DNA-, RNA-, and ChIP-seq data supporting the alternative or the reference allele were counted using Samtools mpileup. Only bases with phred score >20 were considered. Only heterozygous SNPs covered by at least 4 sequencing reads in each data set were included in the final summary.

PCR and Sanger sequencing validation of structural variants. PCR experiments were performed as follows: 10 ng of genomic DNA were used with the SequalPrep Long PCR Kit (Invitrogen) in 20 μl volumes using the following PCR conditions in a MJ Mini thermocycler (BioRad): 94 °C for 3 min, followed by 10 cycles of 94 °C for 10 s, 62 °C for 30 s and 68 °C for 6 min and 25 cycles of 94 °C for 10 s, 60 °C for 30 s and 68 °C for 7 min, followed by a final cycle of 72 °C for 10 min. PCR products were analysed on a 1% agarose gel stained with Sybr Safe Dye (Invitrogen). Gel-extracted bands using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) were capillary sequenced at GATC Biotech AG to analyse SV breakpoints.

Expression array processing and data analysis. *General array processing.* For gene expression array profiling of human medulloblastomas and normal cerebellar controls, high-quality RNAs were processed and hybridized to either (i) the Affymetrix Gene 1.1 ST array at The Centre for Applied Genomics (TCAG, Toronto, Canada) or (ii) the Affymetrix U133 Plus2.0 expression array at the Microarray Department of the University of Amsterdam (Amsterdam, the Netherlands). Sample library preparation, hybridization, and quality control were performed according to protocols recommended by the manufacturer. The CEL files were quantile normalized using Expression Console (v1.1.2; Affymetrix, USA) and signal estimates determined using the RMA algorithm.

Mouse medulloblastomas, non-neoplastic cerebellar stem cells (NSCs), and normal mouse cerebella were analysed using the Affymetrix Mouse Genome 430 2.0 expression array according to the manufacturer's instructions at the DKFZ Genomics and Proteomics Core Facility (Heidelberg, Germany). The CEL files were quantile normalized using Expression Console (v1.1.2; Affymetrix, USA) and signal estimates determined using the RMA algorithm.

Merging of expression array platforms. Gene expression array data generated using the Affymetrix Gene 1.1 ST array and U133 Plus2.0 array platforms was merged in order to generate a combined series that would facilitate more streamlined downstream analyses. For each platform, a contrast value per gene was calculated by subtracting the mean expression of that gene across all samples hybridized on that platform from each individual sample (see formula below), and the resulting contrast values of the two platforms were then combined.

$$\text{Contrast}_{\text{geneA in SampleX}} = \text{Gene}_A \text{ expression in Sample}_X - \text{mean (Gene}_A \text{ expression)}$$

This method minimized possible batch effects existing between the two array platforms and allowed for downstream analyses containing the combined series.

Identification of GFI1- and GFI1B-activated medulloblastomas. After combining the gene expression data for the two expression array platforms, for both *GFI1* and *GFI1B*, expression values were modelled by fitting two normal distributions to the data using the R package 'mclust'[42]. With a *P* value cut-off of $P < 0.0001$, threshold expressions for *GFI1* and *GFI1B* were identified as contrast scores of 0.64 and 0.65, respectively. Samples having expression greater than or equal to the thresholds were called as *GFI1*- or *GFI1B*-activated.

Pathway analysis. Medulloblastoma expression array profiles (Affymetrix Gene 1.1 ST) were used to fit a linear model for each gene using group 3 status, *GFI1* expression, and *GFI1B* expression as covariates. The R package 'limma' was used to perform these fits. The average rank of the statistical significance of the *GFI1* and *GFI1B* coefficients was used to perform a Mann–Whitney *U*-test for a given collection of genes (the null hypothesis being that the genes in a gene set are not ranked any higher than those which are not). In cases where multiple probes matched a single gene, the higher-ranking probe was used. The gene sets contained in the c2-c6 collections from the Molecular Signatures Database (MSigDB) were tested[43]. The *P* values obtained for each gene set in a collection underwent a Benjamini–Hochberg correction to correct for multiple testing.

Cross-species comparisons of human and mouse medulloblastomas. Human medulloblastoma samples were analysed on the Affymetrix U133Plus2 platform

and normalized by the MAS5 algorithm. Mouse tumours were analysed on the Affymetrix Mouse Genome 430 2.0 platform and similarly normalized by MAS5 using the 'affy' (v1.38) package within the R Statistical Environment (v 3.0.2). Human and mouse expression profiles were matched by homologues using official gene symbols and filtered for genes that exhibit conserved expression across 32 matched human and mouse tissues[44] as determined by Pearson correlation tests with multiple hypotheses correction using the Benjamini–Hochberg false discovery rate method (FDR <0.1). Mouse adult cerebellum, fetal cerebellum and *Ptch1*$^{+/-}$ medulloblastoma samples were matched against the most similar human adult cerebellum, fetal cerebellum and SHH medulloblastoma samples, respectively, by Pearson correlation of expression profiles. Subsequently, these matched sample pairs were designated as replicate samples for cross-platform calibration by the Linear Cross-Platform Integration of Microarray Data (LTR) algorithm[45] as implemented in the 'LTR' package (v 1.0.0).

Following gene filtering and expression calibration, the human and mouse expression profiles were combined and analysed by multidimensional scaling. The first two dimensions were disregarded, as expression differences between human and mouse dominated them. The third dimension was identified as the medulloblastoma subgroup spectrum, as the coordinate values discriminate samples from different human medulloblastoma subgroups. Using this molecular subgroup spectrum, mouse samples were classified using a Bayesian classifier initialized with a uniform prior. The posterior probabilities were calculated as the normalized product of the prior and the likelihood of Gaussian distribution parameters with mean and variance estimates from each of the human medulloblastoma subgroups.

Luciferase enhancer assays. Candidate enhancer regions were amplified by PCR using the primer sets listed below and cloned into the pGL4.24[luc2P/minP] Vector (Promega) containing a multiple cloning region for insertion of a response element of interest upstream of a minimal promoter and the luciferase reporter gene, *luc2P*.

Primer sequences. The following primer sequences were used: BARHL1/DDX31 SE: region 1, forward primer GAAGGTACCATCCCCACTTCCTGGTAAGG, reverse primer GAAGGTACCTTCTTTGGGGAAATCATTGG; BARHL1/DDX31 SE: region 2, forward primer GAAGGTACCCTGAGAGTTTGGGCTTCAGG, reverse primer GAAGGTACCGCCTGCCAATTTTTATGTGG; BARHL1/DDX31 SE: region 3, forward primer GAAGGTACCTGTCTCCAAGTGTGGTTTCG, reverse primer GAAGGTACCTGAGCAGGGGATTTAACAGG; BARHL1/DDX31 SE: region 4, forward primer GAAGGTACCAGGGGTATCGTGGTCTTGG, reverse primer GAAGGTACCGGAAAGCACACGTGAAAAGG; BARHL1/DDX31 SE: region 5, forward primer AAGGTACCAGTGTGTCAACCACCCACAA, reverse primer AAGAGCTCGGATGGAGTGCAGTCACCTT; BARHL1/DDX31 SE: region 6, forward primer AAGGTACCGAAATTCCCCAGGAGGAGAG, reverse primer AAGAGCTCCCAATGCACCCTACGTTTCT; BARHL1/DDX31 SE: region 7, forward primer AAGGTACCCACCCAGCTCTTCTCCAGTC, reverse primer AAGAGCTCCTCCTCCAGCACAACACTGA; BARHL1/DDX31 SE: region 8, forward primer AAGGTACCCTGTAGCCTCGACCTTCTGG, reverse primer AAGAGCTCCCTTCAGAGCACTTGTAGGAGAA.

For evaluation of enhancer activity, D425 group 3 medulloblastoma cells were plated on 6-well plates. At 50% confluence, cells were transfected in triplicate with 2.25 μg of the pGL4.24 reporters carrying the *DDX31* DNA fragments plus 0.25 μg of phRL-TK encoding Renilla luciferase. Two days post-transfection, the cells were harvested, followed by measurement of luciferase activities using the Dual-Glo Luciferase Assay System (Promega). As a control, the pGL4.24 empty vector was included for calibration of activity obtained with the experimental constructs. The luminescence of the Firefly Luciferase was normalized to the Renilla Luciferase signal obtained from the phRL-TK vector and data was presented as the mean delta-fold activity (Firefly Luciferase/Renilla Luciferase) of experimental transfectants compared to the pGL4.24 empty vector transfectants.

Immunohistochemical and FISH analysis of human medulloblastoma samples. Immunohistochemistry (IHC) and FISH were performed on formalin-fixed paraffin-embedded MB sections as previously described[7]. Monoclonal GFI1 (clone 3G8, Sigma) and polyclonal GFI1B (HPA007012, Sigma) antibodies were used at working dilutions of 1:100 with an incubation time of 1 h @ 32°C using the Ventana protocol cc1.

Mouse models. *Animals.* C57BL/6 mice (males and females) were used as a source of cerebellar stem cells and immunocompromised (NOD-*scid* IL2Rgamma$^{\text{null}}$, NSG) female mice were used as transplantation hosts. Mice were bred and maintained at the Sanford-Burnham and Sanford Consortium Animal Facilities. Experiments were performed in accordance with national regulations using procedures approved by the Institutional Animal Care and Use Committees at Sanford-Burnham and the University of California San Diego. No a priori calculations related to sample size were performed. No specific randomization or blinding was performed.

Isolation of cerebellar stem cells. Cerebellar stem cells were isolated as previously described[46]. Briefly, neonatal (p4-p6) cerebella from wild-type C57BL/6 mice were dissected and enzymatically dissociated into single cell suspension. Cells were

subjected to Percoll fractionation (GE Healthcare Life Sciences 17-0891-02) and stained (anti-mouse CD133 PE, eBioscience 12-4301-82) and sorted for the Prominin1+ (Prom1$^+$) population (approximately 3–4% of cells).

Retroviral constructs. Retroviruses employed in this study included MSCV-c-Myc^{T58A}-IRES-GFP[21], MSCV-c-Myc T58A-IRES-Luc, MSCV-Gfi1-IRES-GFP, MSCV-Gfi1-IRES-Luc, MSCV-Gfi1b-IRES-GFP and MSCV-Gfi1b-IRES-Luc. To create the *Gfi1* and *Gfi1b* viral constructs, cDNAs were PCR-amplified and cloned into MSCV-IRES-GFP and MSCV-IRES-Luc. *Gfi1* and *Gfi1b* were PCR-amplified from pCMV6-Gfi1 (MC208542, OriGene) and pCMV6-Gfi1b (MC201880, OriGene), respectively, and EcoRI and XhoI restriction sites were added to the cDNA ends.

Gfi1 PCR primers: forward primer GAATTCACCATGCCGCGCTCATTCC TGGTC, reverse primer CTCGAGTCATTTGAGTCCATGCTGACTCTC. *Gfi1b* PCR primers: forward primer GAATTCACCATGCCACGGTCCTTTCTAGTG, reverse primer CTCGAGTCACTTGAGATTGTGTTGACTCTC.

The PCR-amplified products were blunt-end-ligated into pJET1.2 (CloneJET PCR Cloning Kit, Thermo Scientific K1231) and then cut with EcoRI and XhoI. The sticky-ended fragments were then ligated into the EcoRI/XhoI-digested MSCV-IRES-GFP and MSCV-IRES-Luc vectors.

Orthotopic transplantation and tumour formation. Before transplantation, cerebellar stem cells (Prom1$^+$ cells) were infected with retroviruses encoding *Myc*^{T58A} and *Gfi1*- or *Gfi1b* for 20 h. Next, 1×10^5 transduced cells were re-suspended in Neurocult NSC Basal medium (Stem Cell Technologies, cat #05700) with Neurocult NSC Proliferation Supplement (Stem Cell Technologies, cat #05701) and injected into the cerebella of NSG mice (6–8 weeks old) using a stereotactic frame equipped with mouse adaptor (David Kopf Instruments). Animals were monitored weekly and euthanized when they showed signs of medulloblastoma. At time of euthanasia, brains were removed for tumour dissection and dissociation or for embedding and sectioning.

Tissue sectioning and staining. Mouse brains were fixed with 4% paraformaldehyde and embedded in either paraffin or OCT. Samples for histological analysis were paraffin-embedded, sectioned and stained with H&E by the Sanford-Burnham Histopathology Core Facility. Samples frozen in OCT were sectioned using a Leica CM3050S cryostat. Cryosections were stained overnight with primary antibodies against proliferation (anti-Ki67, Abcam ab15580) and lineage markers (anti-GFAP, Novus Biologicals NB300-141; anti-β3-Tubulin, Cell Signaling 5568) and stained for 1 h with fluorescent secondary antibodies (Alexa Fluor 568 Donkey Anti-Rabbit IgG, Invitrogen A10042). Sections were then counter-stained with DAPI (Cell Signaling 4083), mounted using Fluoromount G (Southern Biotech #0100-01), and imaged on a confocal (Zeiss LSM700) fluorescent microscope.

In vivo bioluminescent imaging. Mice were anaesthetized with 2.5% isoflurane and given intraperitoneal injections of 150 ng g^{-1} D-Luciferin (Caliper Life Sciences, cat 12279). Five minutes after injection, animals were imaged using the Xenogen Spectrum (IVIS-200) imaging system (Sanford-Burnham and Sanford Consortium Animal Facilities).

34. Richter, J. *et al.* Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nature Genet.* **44,** 1316–1320 (2012).
35. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).
36. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28,** 333–339 (2012).
37. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
38. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
39. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
40. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).
41. Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12,** R72 (2011).
42. Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E. & Ruzzo, W. L. Model-based clustering and data transformations for gene expression data. *Bioinformatics* **17,** 977–987 (2001).
43. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
44. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA* **101,** 606–6067 (2004).
45. Boutros, P. C. LTR: Linear cross-platform integration of microarray data. *Cancer Inform.* **9,** 197–208 (2010).
46. Lee, A. *et al.* Isolation of neural stem cells from the postnatal cerebellum. *Nature Neurosci.* **8,** 723–729 (2005).
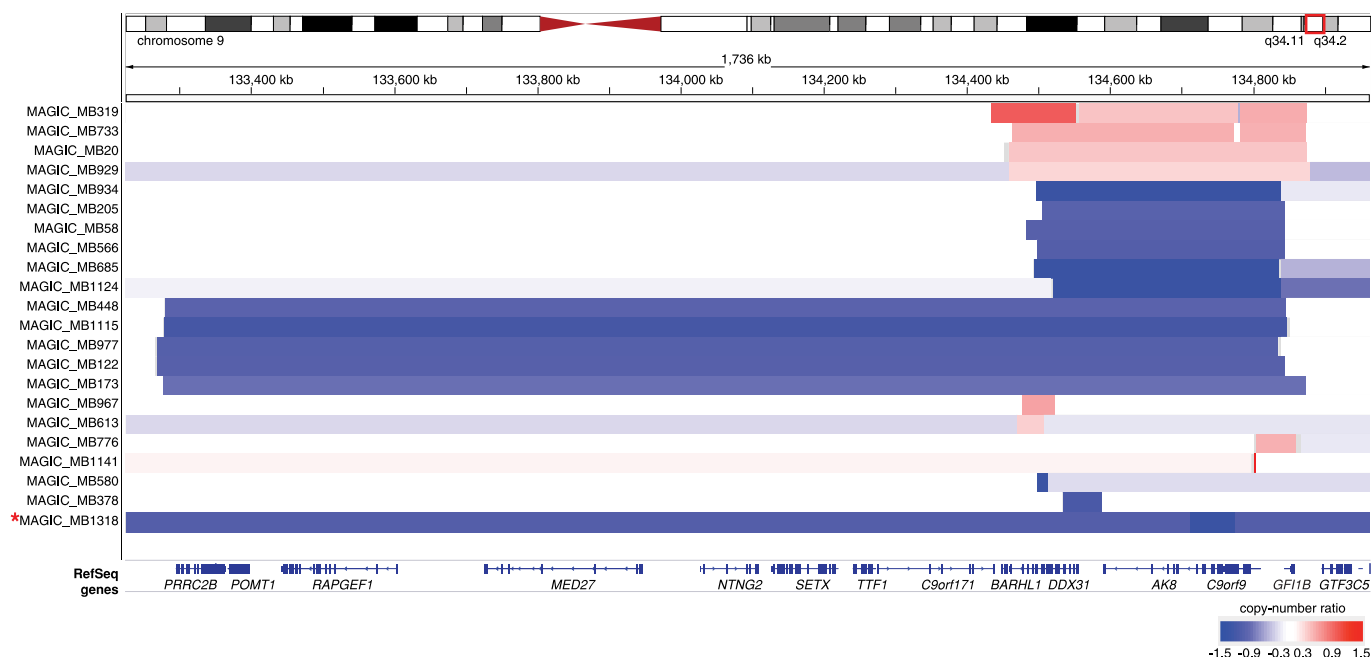
Extended Data Figure 1 | Recurrent somatic copy-number aberrations target a common region on 9q34. Affymetrix SNP6 copy-number output for 22 primary medulloblastomas from the published[8] MAGIC series exhibiting focal somatic copy-number aberrations within the 9q34 region of interest defined by WGS in the current study. Of the affected samples, medulloblastoma subgroup information was available for 15 of 22 cases: SHH (*n* = 1*), group 3 (*n* = 11) and group 4 (*n* = 3). Close examination of the single non-group 3/group 4 medulloblastoma affected by a focal copy-number event in the region (MAGIC_MB1318, SHH) revealed that this sample exhibits a homozygous deletion (in the context of broad chr9q deletion) specifically overlapping *TSC1* and is therefore unlikely to be related to the events which target *GFI1B* for transcriptional activation. Indicated coordinates are based on the hg18 reference genome (NCBI Build 36.1) that was used in the original MAGIC study.

a

b
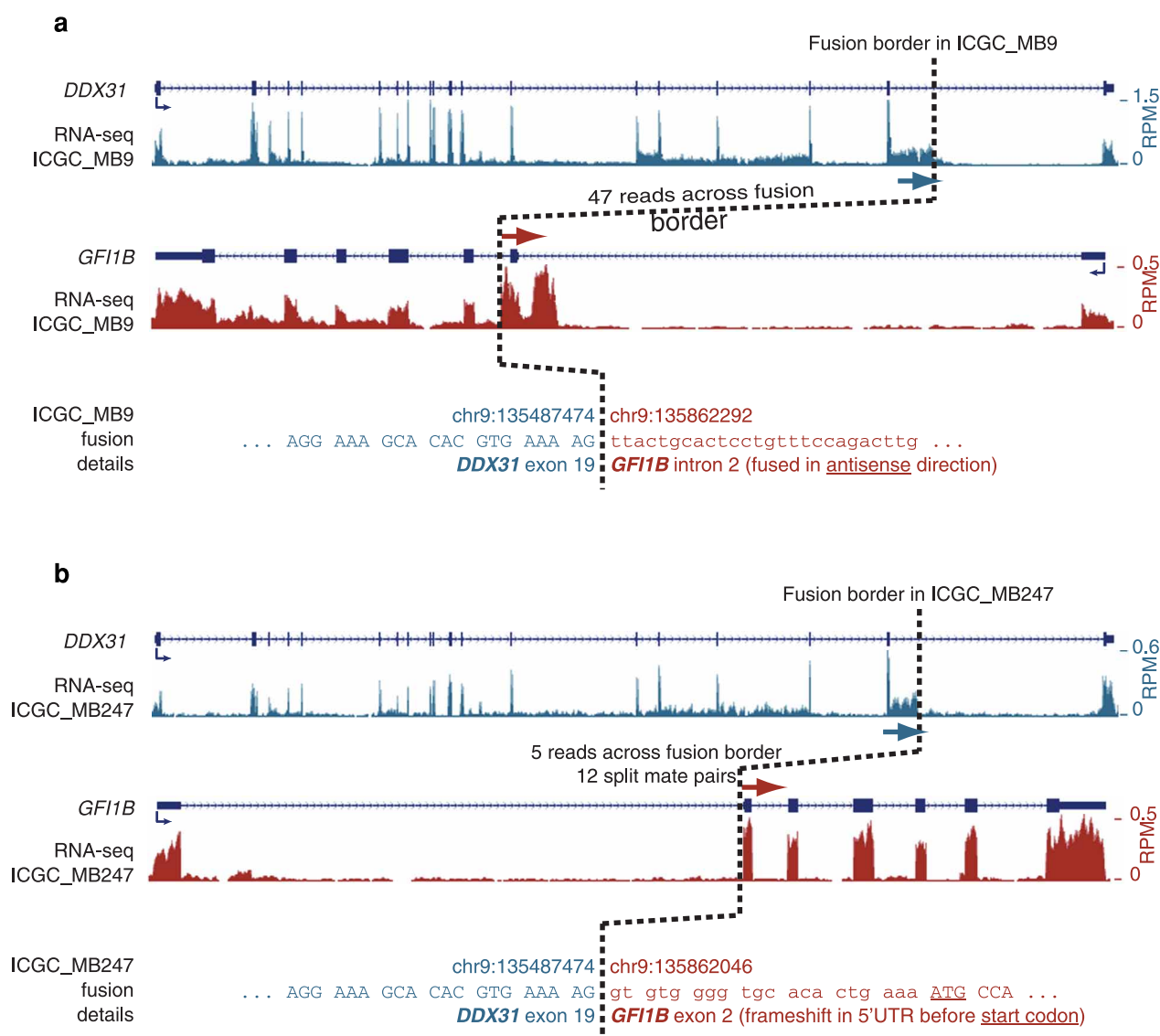
Extended Data Figure 2 | Non-functional *DDX31–GFI1B* fusion transcripts detected by RNA-seq. a, A complex SV on 9q34 in ICGC_MB9 resulted in expression of *DDX31* (exon 19) fused to *GFI1B* (intron 2, antisense orientation). Note the intronic reads in *GFI1B* after the fusion breakpoint. **b,** 9q34 inversions in ICGC_MB247 resulted in expression of *DDX31* (exon 19) fused to *GFI1B* (exon 2, sense orientation). This fusion transcript included a frameshift, inferred to generate a C-terminal-truncated DDX31 protein and no GFI1B protein from this fused allele.
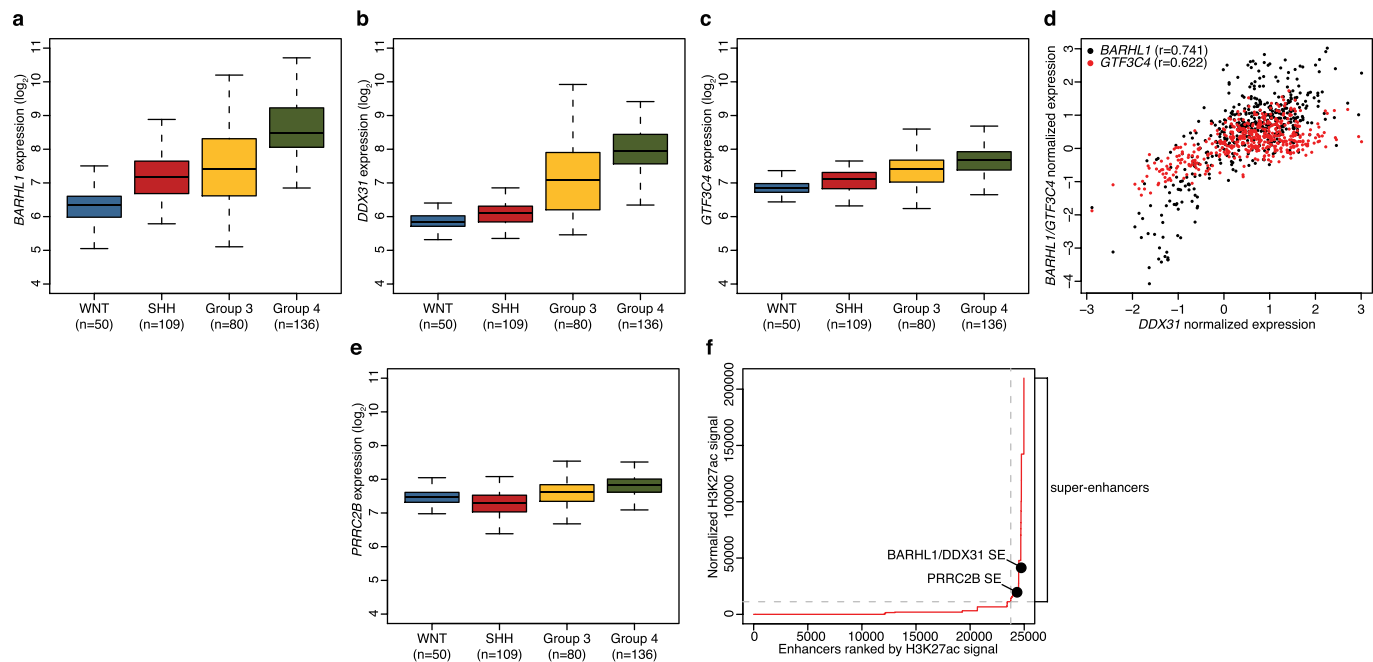
Extended Data Figure 3 | Expression and correlation of 9q34 genes in medulloblastoma subgroups. a–c, Box-plots summarizing expression of *BARHL1* (**a**), *DDX31* (**b**) and *GTF3C4* (**c**) according to medulloblastoma subgroup. Data set includes 375 medulloblastomas profiled on the Affymetrix U133plus2 array. **d,** Pearson correlation analysis showing correlated expression of *DDX31* with *BARHL1* and *GTF3C4* in group 3 and group 4 medulloblastomas. *DDX31* expression is positively correlated with both *BARHL1* ($r = 0.741$) and *GTF3C4* ($r = 0.622$). **e,** *PRRC2B* expression in

medulloblastoma subgroups. Samples are from the same series summarized in **a–c**. **f,** Distribution of H3K27ac ChIP-seq signal at predicted enhancers in group 3 medulloblastomas (data for MAGIC_MB360 are shown). Enhancer regions are plotted in increasing order based on their input-normalized H3K27ac signal. Super-enhancers are defined as the population of enhancers above the inflection point of the curve (horizontal dashed grey line). Positions of the predicted *BARHL1/DDX31* and *PRRC2B* super-enhancers described in the text are highlighted.
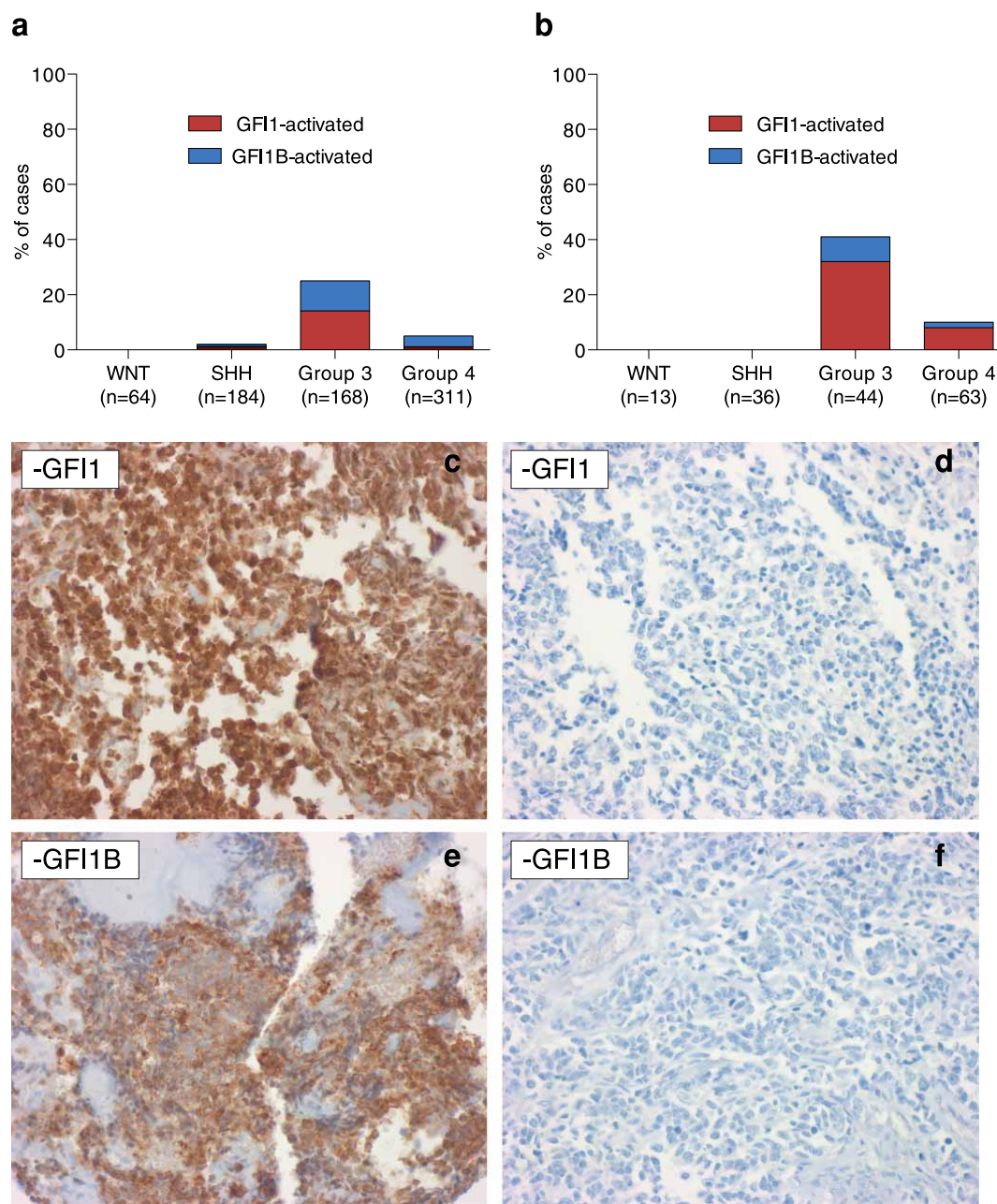
Extended Data Figure 4 | Frequency and distribution of GFI1/GFI1B activation in medulloblastoma subgroups. **a**, Stacked bar graph indicates the proportion of *GFI1/GFI1B*-expressing cases in each of the four medulloblastoma subgroups, as determined by Affymetrix gene expression profiling of two independent cohorts ($n = 727$). **b**, Stacked bar graph indicates the proportion of GFI1/GFI1B-positive cases in each of the four medulloblastoma subgroups, as determined by immunohistochemistry performed with anti-GFI1 and anti-GFI1B antibodies on formalin-fixed paraffin-embedded sections derived from a medulloblastoma clinical trial cohort (HIT2000, NCT00303810; $n = 156$). **c–f**, Representative positive and negative immunohistochemistry results for group 3 medulloblastomas stained with anti-GFI1 (**c**, **d**) and anti-GFI1B (**e**, **f**) antibodies, respectively.
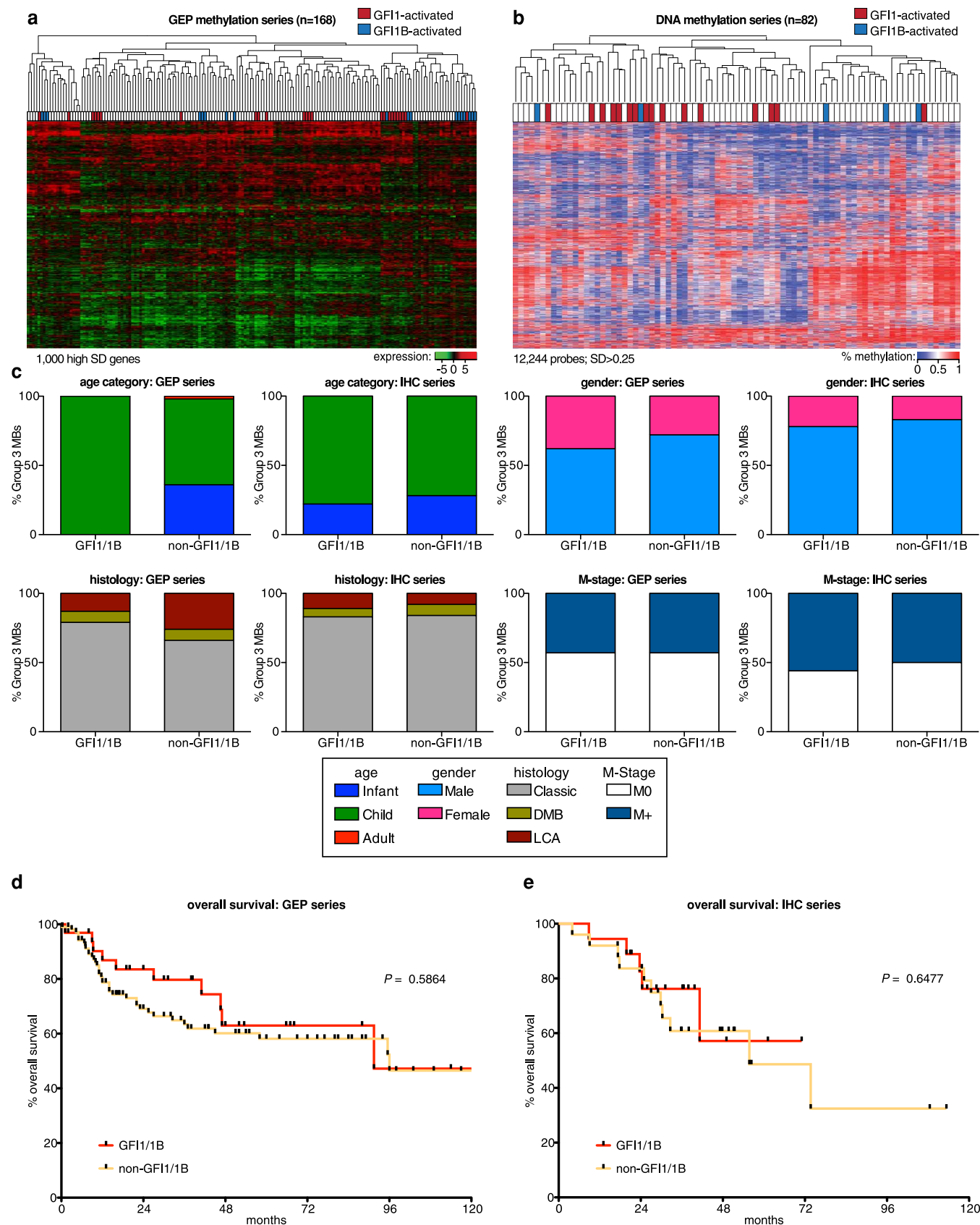
Extended Data Figure 5 | Demographic and clinical characteristics of GFI1/GFI1B-activated group 3 medulloblastoma. a, b, Unsupervised hierarchical clustering of group 3 medulloblastoma samples profiled by Affymetrix gene expression array (**a**) or Illumina 450K DNA methylation array (**b**). **c,** Patient characteristics, including age, gender, histological subtype (histology) and metastatic status (M-stage) for group 3 medulloblastomas stratified according to GFI1 and GFI1B expression status. Both gene expression and immunohistochemistry cohorts are summarized. **d, e,** Overall survival of group 3 medulloblastomas stratified by GFI1 and GFI1B expression status for both our gene expression (**d**) and immunohistochemistry series (**e**).
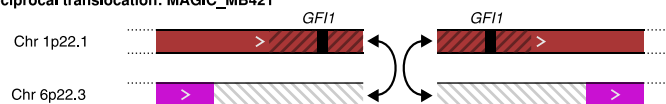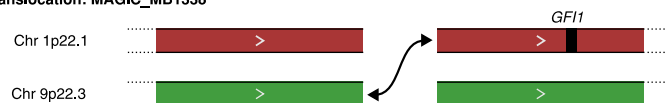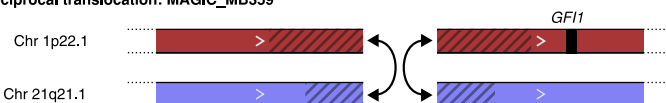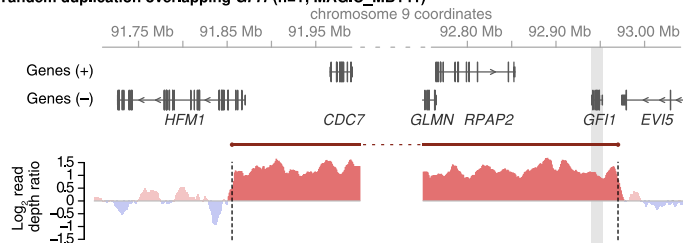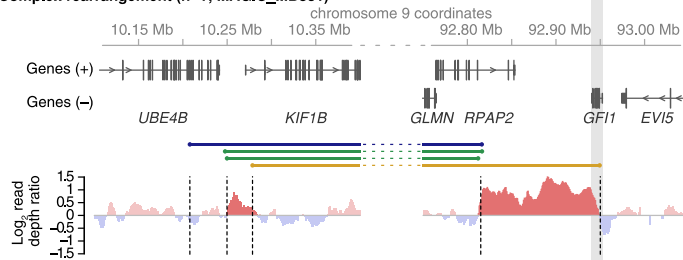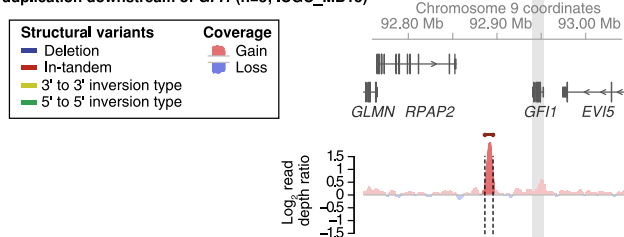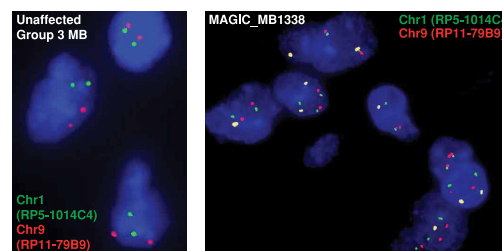
Extended Data Figure 6 | Summary of *GFI1* SVs detected by WGS in group 3 medulloblastoma. a, Schematics depicting the six different *GFI1* translocations detected by large-insert paired-end sequencing of our *GFI1*-activated validation series. **b**, WGS coverage plots showing SVs affecting the *GFI1* locus in *GFI1*-activated medulloblastomas sequenced in our series. **c**, Fluorescence *in situ* hybridization (FISH) analysis of MAGIC_MB1338 validating the unbalanced t(1;9) translocation (shown in **a**) predicted by WGS.
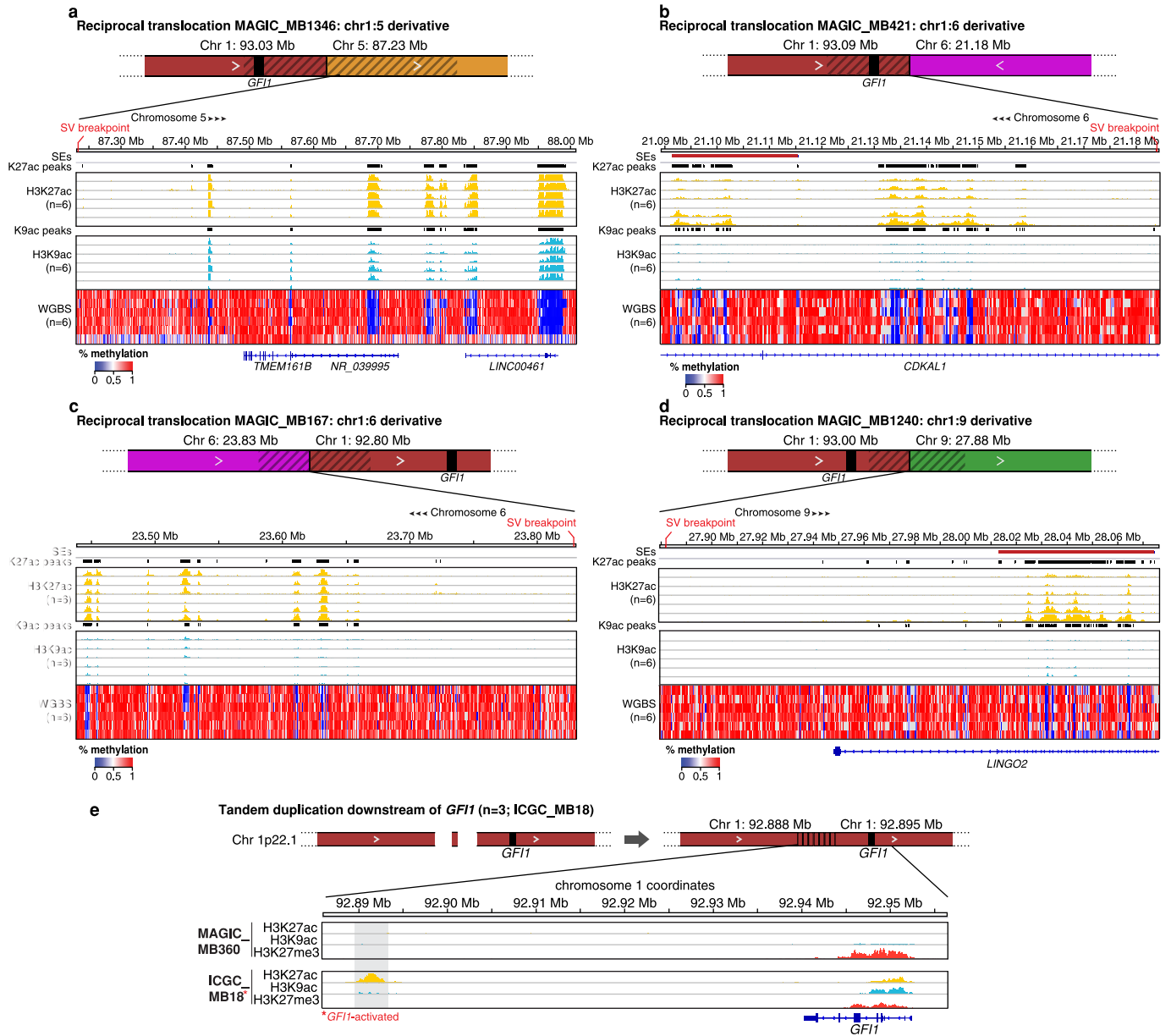
Extended Data Figure 7 | Chromatin states proximal to SVs observed in *GFI1*-activated group 3 medulloblastomas. a–d, ChIP-seq (H3K27ac and H3K9ac) and WGBS data respectively highlighting the active chromatin and methylation states present in the regions proximal to SV breakpoints identified in *GFI1* translocation cases. **e**, Schematic summarizing the series of focal tandem duplications observed approximately ~45 kb downstream of *GFI1* in group 3 medulloblastomas (*n* = 3; ICGC_MB18 is shown as a representative case). Activating and repressive histone marks overlapping the region of interest are shown for a non-*GFI1*-activated group 3 medulloblastoma (MAGIC_MB360) and the tandem duplication case (ICGC_MB18).

a

b

c2 (CGP, Biocarta, KEGG, Reactome)

gene set	# of observed genes	FDR (0.25)
CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_2	430	<<0.0001
KIM_MYC_AMPLIFICATION_TARGETS_UP	187	0.00159
VERHAAK_GLIOBLASTOMA_CLASSICAL	196	0.0123

c

d

'KIM_MYC_AMPLIFICATION_TARGETS_UP' gene set
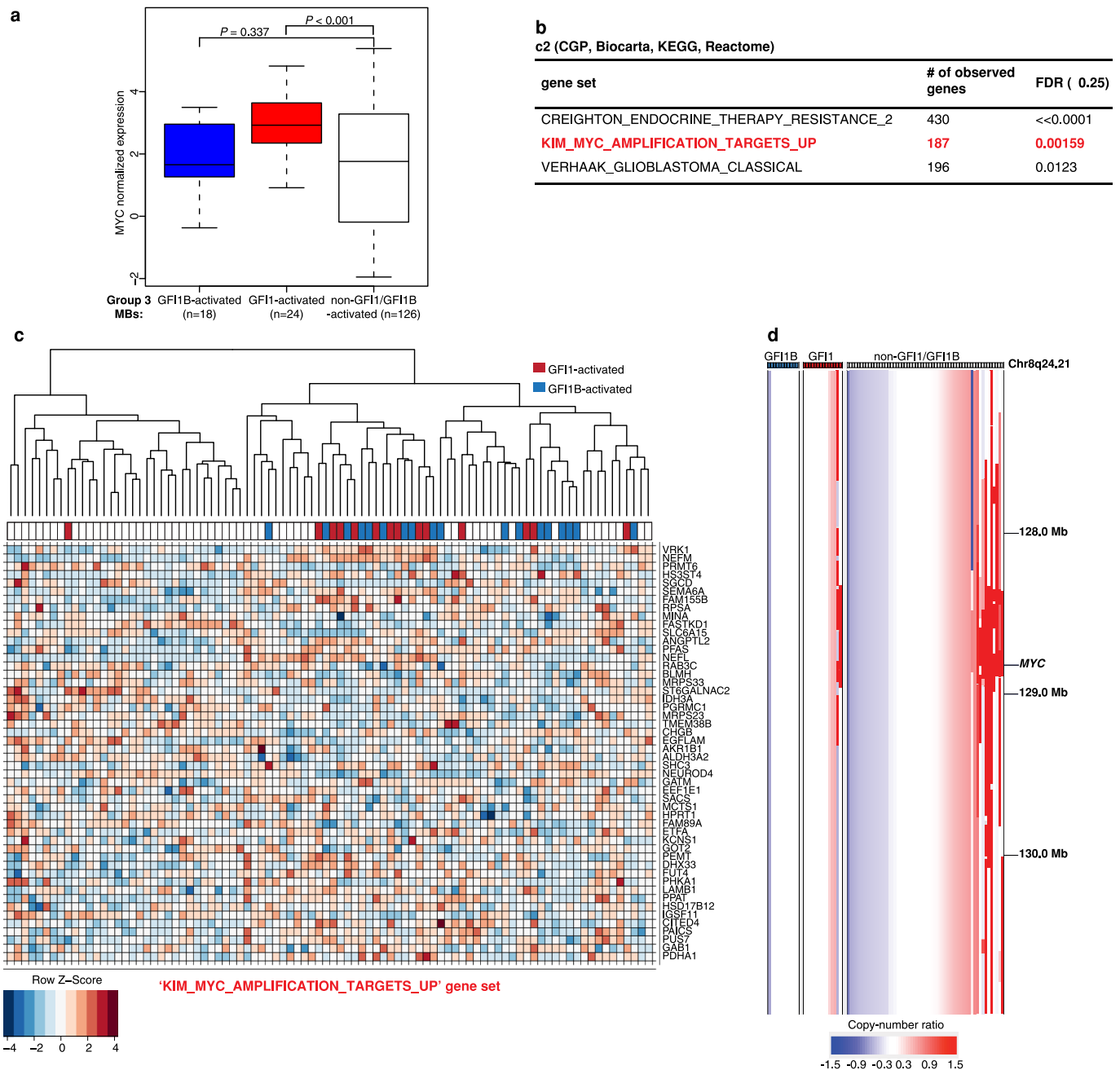
Extended Data Figure 8 | Association between *GFI1/GFI1B* activation and *MYC* in group 3 medulloblastoma. a, *MYC* expression in group 3 medulloblastomas (*n* = 168) according to *GFI1* and *GFI1B* activation status. **b**, Gene sets with significant enrichment in *GFI1/GFI1B*-associated genes from the MSigDB c2 gene set collection. The collection highlighted in red is the only result found that shows a significant enrichment in both *GFI1* and *GFI1B* associated genes and a clear connection to a known pathway. **c**, Heat-map of the expression values for the 50 genes in the KIM_MYC_AMPLIFICATION_TARGETS_UP gene set with the most significant association with *GFI1* or *GFI1B* expression (the complete gene set contains 187 profiled genes). Genes are ordered top to bottom from most to least significant. A set of 90 group 3 medulloblastomas included in the analysis is displayed. Sample-wise hierarchical clustering was performed only to enhance the visual organization of the heat map. **d**, Affymetrix SNP6 copy-number output for 82 primary group 3 medulloblastomas from the published MAGIC series, highlighting the incidence of *MYC* amplification in the context of *GFI1/GFI1B*-activation. *MYC* amplification was found at a comparable frequency in both *GFI1*-activated (*n* = 2 of 14, 14.3%) and non-*GFI1/GFI1B*-activated (*n* = 10 of 57, 17.5%) group 3 medulloblastomas. Indicated coordinates are based on the hg18 (NCBI Build 36.1) reference genome that was used in the original MAGIC study.
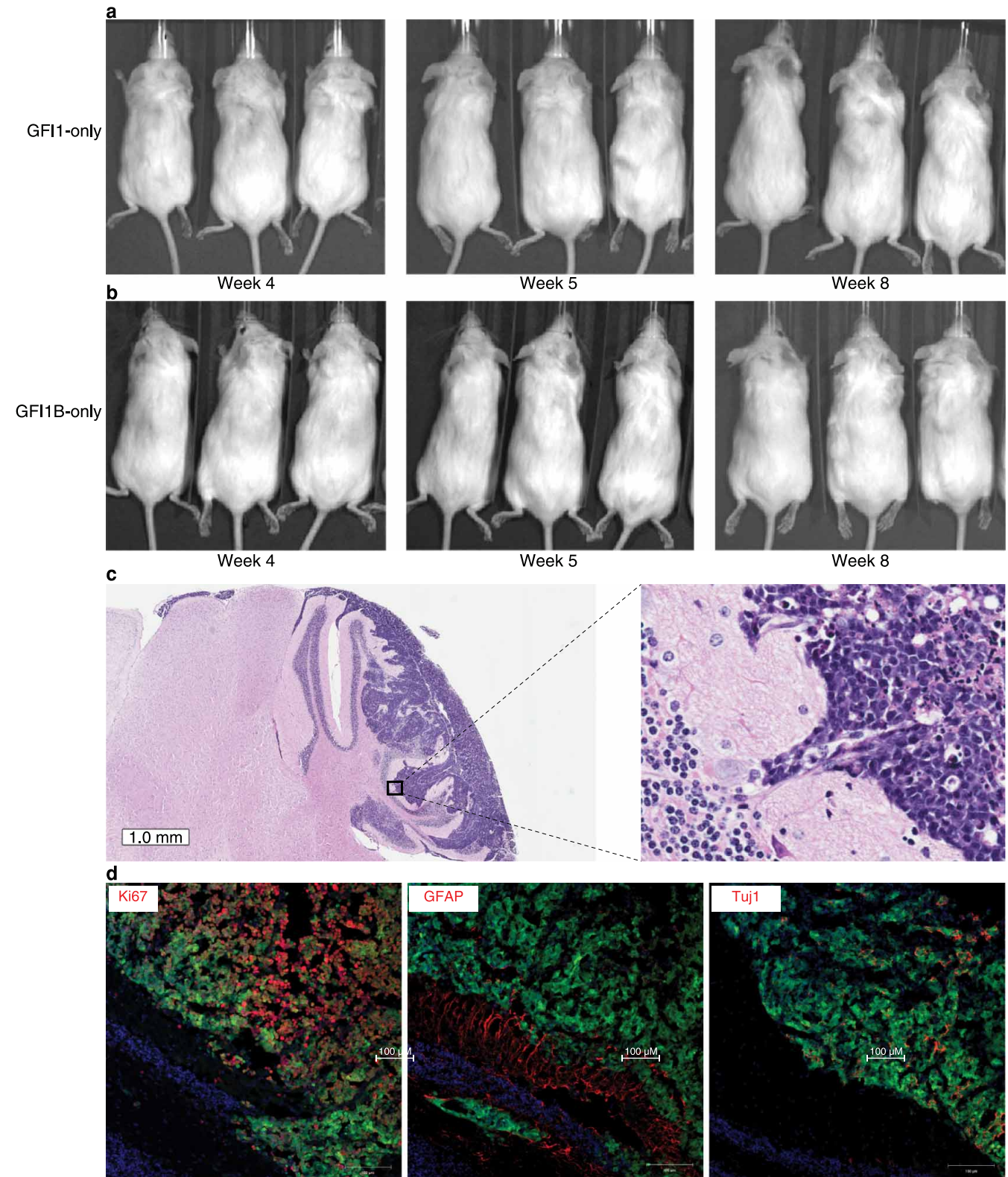
Extended Data Figure 9 | Phenotypic characteristics of novel *GFI1/GFI1B* orthotopic mouse models. **a**, **b**, Bioluminescent imaging of animals injected with either *GFI1-* (**a**) or *GFI1B*-expressing (**b**) neural stem cells at the indicated time points. No tumour signal was detectable in these animals.

c, Haematoxylin and eosin staining of cerebellar sections derived from MYC + GFI1B tumour-bearing mice. **d**, Immunofluorescence imaging of cerebellar sections from MYC + GFI1B tumours stained with the indicated antibodies.

ARTICLE

Structure of an Rrp6–RNA exosome complex bound to poly(A) RNA

Elizabeth V. Wasmuth[1,2], Kurt Januszyk[1] & Christopher D. Lima[1,3]

The eukaryotic RNA exosome processes and degrades RNA by directing substrates to the distributive or processive 3′ to 5′ exoribonuclease activities of Rrp6 or Rrp44, respectively. The non-catalytic nine-subunit exosome core (Exo9) features a prominent central channel. Although RNA can pass through the channel to engage Rrp44, it is not clear how RNA is directed to Rrp6 or whether Rrp6 uses the central channel. Here we report a 3.3 Å crystal structure of a ten-subunit RNA exosome complex from *Saccharomyces cerevisiae* composed of the Exo9 core and Rrp6 bound to single-stranded poly(A) RNA. The Rrp6 catalytic domain rests on top of the Exo9 S1/KH ring above the central channel, the RNA 3′ end is anchored in the Rrp6 active site, and the remaining RNA traverses the S1/KH ring in an opposite orientation to that observed in a structure of a Rrp44-containing exosome complex. Solution studies with human and yeast RNA exosome complexes suggest that the RNA path to Rrp6 is conserved and dependent on the integrity of the S1/KH ring. Although path selection to Rrp6 or Rrp44 is stochastic *in vitro*, the fate of a particular RNA may be determined *in vivo* by the manner in which cofactors present RNA to the RNA exosome.

The eukaryotic RNA exosome core is an essential nine-subunit complex formed by two stacked rings of six RNase PH-like proteins on the bottom and three S1/KH domain 'cap' proteins on the top. Exo9 associates with the endo- and 3′ to 5′ processive exoribonuclease Rrp44, and the 3′ to 5′ distributive exoribonuclease Rrp6 (ref. 1). In *Saccharomyces cerevisiae*, the cytoplasmic exosome includes Exo9 and Rrp44 (Exo10^{Rrp44}), the nuclear exosome contains Exo9, Rrp44 and Rrp6 (Exo11$^{Rrp44/Rrp6}$)[2], and a nucleolar exosome with Exo9 and Rrp6 (Exo10^{Rrp6}) has been posited in human[3]. Each exosome complex may be uniquely equipped to target different RNA substrates, in conjunction with protein cofactors, to catalyse RNA turnover, quality control, or processing in the context of their respective cellular compartments[4].

Rrp6 is a RNase D family member[5], and is proposed to hydrolyse RNA via two metal ion catalysis[6,7]. It includes an amino-terminal PMC2NT domain that associates with a cofactor Rrp47 (ref. 8), an exoribonuclease domain (EXO), an HRDC domain, and a carboxy-terminal domain (CTD) that associates with the Exo9 core[9]. Structural studies revealed that the EXO and HRDC domains constitute the catalytic module[10,11] and a Exo10$^{Rrp44+Rrp6Cterm}$ structure revealed how Rrp6 CTD residues interact with Exo9 (ref. 12). No structures yet exist for Rrp6 in complex with Exo9 or RNA.

We proposed that Rrp44 and Rrp6 use an overlapping channel within the S1/KH ring to engage RNA, that Rrp6 activities are modulated by Exo9, and that Rrp6 stimulates Rrp44 in binding and degradation of single stranded RNA in the context of Exo11$^{Rrp44/Rrp6}$ (ref. 13). To provide a structural basis for these observations, we crystallized Rrp6 with Exo9 and a 24-nucleotide single-stranded poly(A) RNA (poly(A)$_{24}$ RNA).

Global architecture of Exo10^{Rrp6}–poly(A) RNA

The 3.3 Å structure of Exo10^{Rrp6} was obtained in the presence of a 24-nucleotide (nt) single-stranded poly(A) RNA using Rrp6 (128–685) that lacked exoribonuclease activity (D238N)[13,14], the PMC2NT domain (1–127)[13] and its last 48 C-terminal residues (Extended Data Fig. 1; Extended Data Table 1). Rrp6 is positioned atop the Exo9 S1/KH ring with Exo9 subunits resembling those of human Exo9 (ref. 15) and yeast

Exo9 (ref. 12) in Exo10$^{Rrp44+Rrp6Cterm}$ (Fig. 1a, b). Poly(A)$_{24}$ RNA is coordinated within the S1/KH ring with Rrp6 active site residues contacting the RNA 3′ end (Fig. 1c). The Rrp6 EXO domain contacts two of the three S1/KH ring proteins, Rrp4 and Rrp40, and the HRDC domain is proximal to Rrp4, although no direct contacts are apparent. Consistent with previous results[12], the Rrp6 CTD wraps around the N-terminal domain (NTD) of Csl4 and PH-like subunit, Rrp43, before emerging at the top of the Exo9 core (Fig. 1a; Extended Data Fig. 2a, b). No electron density was observed for CTD residues 629–684 and for residues (517–524) that link the HRDC domain and CTD. The location of the Rrp6 N terminus (Met 128) places the Rrp6 PMC2NT domain over the Exo9 central channel in an ideal position to interact with Rrp47 to facilitate substrate recruitment (Fig. 1a). Alignment of exosome-associated Rrp6 to the yeast Rrp6 catalytic module (EXO-HRDC) reveals few differences (Extended Data Fig. 2c).

The NTD of Rrp4 and S1 domains of Rrp4 and Rrp40 interact with the Rrp6 EXO domain, burying 2,230 Å^2 of surface area in the complex. Rrp6 interactions with the Rrp4 NTD (Region 2) and S1 domain (Region 3) are more extensive (1,750 Å^2), whereas interactions with the Rrp40 S1 domain and Rrp4 S1 domain encompass a region proximal to the RNA binding site (Region 1). These surfaces are highly conserved (Fig. 2a, b) in comparison to the Rrp6 CTD (Extended Data Fig. 2a), which is important for Rrp6 interaction with the exosome core as illustrated by analytical gel-filtration studies showing that the CTDs of both yeast and human Rrp6 are required for association with the exosome (Extended Data Fig. 3).

The structure of Exo10^{Rrp6} bound to poly(A)$_{24}$ RNA reveals how an RNase D family member interacts with RNA (Fig. 2c). The Rrp6 catalytic domain and active site bind A$_{21}$A$_{22}$A$_{23}$A$_{24}$ with the 2′-OH and 3′-OH of the terminal A$_{24}$ coordinated via two main-chain hydrogen bonds to the backbone amide and carbonyl oxygen of His 241 and Glu 240 side-chain carboxylate, similar to that observed for AMP in the Rrp6 catalytic domain[10]. An additional contact is observed to A$_{24}$ between the N7 adenine atom and Gln 345 side chain amide. The scissile phosphate of A$_{24}$ and A$_{23}$ is coordinated by a single magnesium ion that bridges

[1]Structural Biology Program, Sloan-Kettering Institute, 1275 York Avenue, New York, New York 10065, USA. [2]Louis V. Gerstner Jr. Graduate School of Biomedical Sciences, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [3]Howard Hughes Medical Institute, 1275 York Avenue, New York, New York 10065, USA.
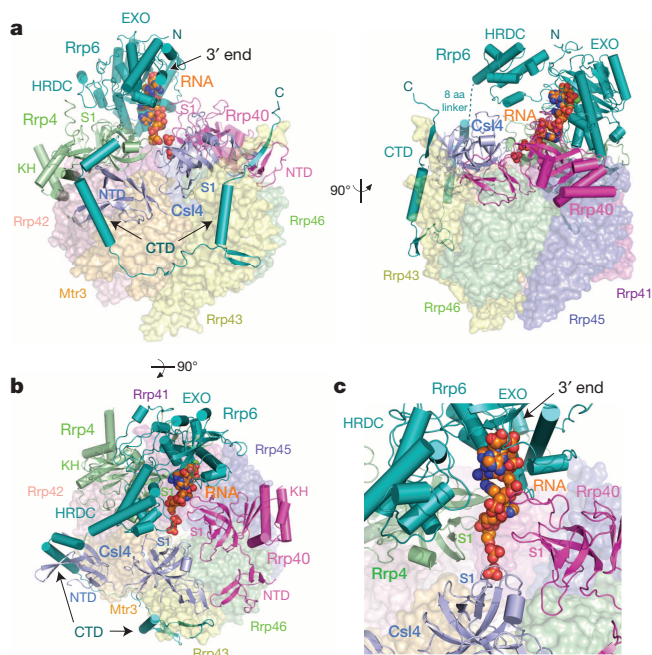
Figure 1 | Overall structure of a RNA bound Rrp6 exosome. a–c, Cartoon and surface representation of *S. cerevisiae* Exo10^{Rrp6} bound to poly(A)$_{24}$ RNA. Views from the sides (**a**), top (**b**) and close-up (**c**) of RNA engaged by the Rrp6 EXO domain and S1 domains of Rrp40, Rrp4 and Csl4. RNase PH-like ring as transparent surfaces, S1/KH ring and Rrp6 as cartoons, RNA as spheres.
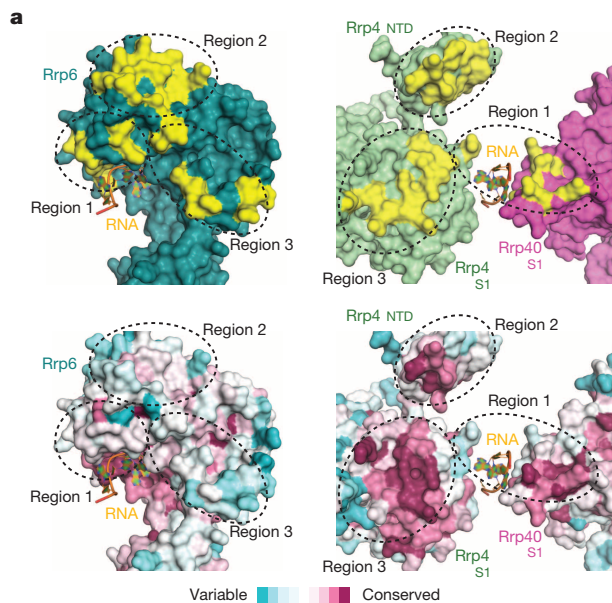
the phosphate and the side chain carbonyl of Asn 238 (Asp in wild-type enzyme). The scissile phosphate is within hydrogen bond distance of the invariant side chain hydroxyl of Tyr 361 (Extended Data Fig. 4a,c). Consistent with the proposed reaction mechanism[7], two metal ions coordinated by conserved side-chain carboxylates in the yeast Rrp6–AMP complex are ideally positioned to contact the scissile phosphate (Extended Data Fig. 4c, d). Several contacts between metal ions and RNA are missing in our structure owing to the D238N substitution.

The remaining contacts to poly(A)$_{24}$ RNA reveal several non-specific interactions, consistent with the ability of Rrp6 to degrade RNA of any sequence. A$_{23}$ 2′-OH is within hydrogen-bonding distance to the side-chain carboxylate of Asp 296, whereas the A$_{23}$ phosphate is within hydrogen-bonding distance of the Leu 328 backbone amide. The A$_{23}$ base is situated over the A$_{24}$ base with van der Waals contacts provided by side chains of Trp 299, Met 295 and Phe 294. Phe 294 is positioned between the A$_{23}$ and A$_{22}$ bases, presumably disrupting base-stacking interactions. The A$_{22}$ base and ribose are cradled by van der Waals contacts to Phe 294, Gly 292, His 291 and Tyr 315 while its phosphate is within hydrogen bonding distance with the backbone amide of His 291. Lys 319, His 326 and Tyr 315 contribute the remaining contacts to the A$_{21}$ base. Importantly, none of the aforementioned contacts seem to interrogate the identity of the base. Sixteen of eighteen residues observed in direct contact to the RNA or metal cofactors are conserved in human RRP6 (Extended Data Fig. 4a, b)[10,11].

RNA passes through the S1/KH ring

Four out of six nucleotides of poly(A)$_{24}$ RNA observed in electron density (5′-A$_{19}$-A$_{24}$-3′) are coordinated by Rrp6 placing the 5′ end (A$_{19}$) proximal to conserved, basic side chains from Rrp4 (Lys 122 and Arg 123),

Figure 2 | Rrp6 contacts the S1/KH ring and RNA. a, b, Rrp6 EXO domain interface with Rrp4 and Rrp40 as yellow buried surfaces (**a**) and in atomic detail (**b**). **a**, Top, Rrp6 (teal; left), Rrp4 (pale green; right) and Rrp40 (magenta; right) in open-book representation (160°). Bottom, surfaces coloured by sequence conservation from red (conserved) to blue (variable). **b**, Contacts between Rrp6, Rrp4 and Rrp40 in stick representation coloured as in **a**. **c**, Contacts between RNA (orange), magnesium (green) and Rrp6 active site residues (teal) in Exo10$^{Rrp6-RNA}$. A catalytic aspartate, D238, is mutated (D238N; red). Potential hydrogen bonds depicted as dashed lines.
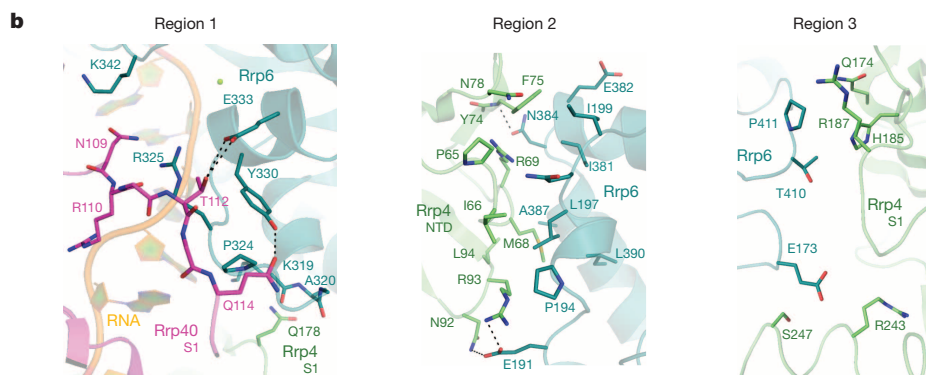
while Rrp40 Lys 108 points towards A_{24} and Arg 110 is proximal to the phosphate of A_{21} and ribose of A_{20} (Fig. 3a). Additional densities with spacing consistent with the RNA phosphate backbone, modelled as PO_{4a} and PO_{4b}, are observed within the S1/KH ring nearest to Arg 150 and proximal to Arg 145 and Arg 202 of Csl4 (Extended Data Fig. 1a–c). The aforementioned Csl4, Rrp4 and Rrp40 residues are highly conserved[15]. No additional density consistent with RNA is observed within the PH-like ring.

Rrp6 and Rrp44 use a shared portion of the central channel to engage RNA substrates[13], and alignment of Exo10[Rrp6] and Exo10[Rrp44 + Rrp6Cterm] reveals overlapping paths for RNA within the S1/KH central channel (Fig. 3a, b), albeit in opposing directions. A subset of side chains from the S1/KH cap proteins contact RNA in both structures, but none make contacts that would enforce directionality. To determine the importance of the S1/KH ring for RNA degradation, yeast Exo10[Rrp6] and Exo11[Rrp44/Rrp6] were reconstituted with Csl4 and Rrp40 containing amino acid substitutions for Arg 145, Arg 150 and Arg 202 in Csl4, and for Lys 107, Lys 108 and Arg 110 in Rrp40. These mutations severely diminish Rrp6 activity in Exo10[Rrp6] and Exo11[Rrp44/Rrp6] as well as Rrp44

activity in Exo11[Rrp44/Rrp6] (Fig. 3c; Extended Data Fig. 5a). In contrast, Rrp6 activity is not inhibited by an insertion in Rrp45 that occludes the PH-like ring channel below the S1/KH ring[13], although this mutation inhibits Rrp44 activity. Reconstituted human Exo10[RRP6] (Extended Data Fig. 3b) reveals similar dependencies as a channel-occluding insertion in RRP41 near the S1/KH ring impaired RRP6 activity, whereas a channel-occluding insertion in RRP45 well below the S1/KH ring failed to inhibit RRP6 (Fig. 3d; Extended Data Fig. 5b). Combining insertions inhibits RRP6 to a greater degree, probably because the RRP45 insertion restricts movement of the RRP41 insertion. These data indicate that the RNA path to Rrp6 depends on the integrity of the S1/KH ring, but does not require the full extent of the PH-like ring central channel. In contrast, at least for yeast Exo11[Rrp44/Rrp6], the RNA path to Rrp44 relies on the integrity of central channel throughout both S1/KH and PH-like rings. These data are consistent with contacts observed to RNA in structures of Exo10[Rrp6] and Exo10[Rrp44 + Rrp6Cterm].

To map RNA interactions through Exo9 to Rrp6 and Rrp44, catalytically dead yeast exosome complexes were reconstituted and subjected to long-wavelength ultraviolet crosslinking using 36-nt poly(A) or AU-rich RNA substrates bearing single 4-thioU substitutions 6, 21 or 29 nt from the 3′ end (Fig. 4a, Extended Data Fig. 6a–c). As the 4-thioU probe is moved 3′ to 5′, AU-rich RNA crosslinks to Rrp44, followed by the PH-like ring, and finally to the S1/KH cap in Exo10[Rrp44] and Exo11[Rrp44/Rrp6], in agreement with previous studies[12,13,16,17]. In contrast, crosslinks are only observed to Rrp44 in Exo10[Rrp44] with poly(A)$_{36}$ RNA; however, addition of Rrp6 in Exo11[Rrp44/Rrp6] results in crosslinks that progress from Rrp44 through the PH-like ring to the S1/KH ring and Rrp6 as the 4-thioU probe is moved 3′ to 5′. The results with poly(A)$_{36}$ RNA are in line with previous biochemistry showing that Rrp6 stimulates Rrp44 binding to poly(A) RNA in Exo11[Rrp44/Rrp6] (ref. 13).

Crosslinking to the Exo10[Rrp6] complex reveals RNA contacts to Rrp6 and Rrp4 when 4-thioU is 6 nt from the 3′ end. Although no structural impediment exists to prevent RNA from entering the PH-like ring, poly(A)$_{36}$ and AU-rich RNA crosslinks are only observed to Rrp4, Csl4 and Rrp40 even when the 4-thioU is positioned 21 or 29 nt from the 3′ end (Fig. 4c; Extended Data Fig. 6). This pattern contrasts with that observed for Exo10[Rrp44] and Exo11[Rrp44/Rrp6], where crosslinks are observed to PH-like ring subunits when 4-thioU is placed 21 nt from the 3′ end. Crosslinking patterns to human Exo10[RRP6] used a 36-nt AU-rich substrate as reported previously[13]. Similar to results obtained for yeast, ultraviolet-induced crosslinks are observed to the three S1/KH ring proteins and RRP45, but to none of the other five PH-like proteins (Fig. 4b). Ultraviolet crosslinking to complexes with loop insertions mirror results observed in decay assays (Fig. 3d) with diminished crosslinking when a loop insertion is placed proximal to the S1/KH ring (Extended Data Fig. 7). These data indicate that the integrity of the S1/KH ring central channel is important for Rrp6 activity in both human and yeast systems.

The results are consistent with a model in which distinct but overlapping paths guide RNA to Rrp6 or to Rrp44 (Fig. 4c). So how is a path selected? Path selection seems stochastic *in vitro* because degradation products of Rrp6 and Rrp44 are observed under conditions of limiting enzyme (Fig. 3c) or limiting substrate (Fig. 4d). The distributive mechanism underlying Rrp6 activity suggests repeated substrate binding and release, whereas the processive mechanism used by Rrp44 suggests that it binds and holds onto substrates until completely degraded. Thus at steady state, binding and ultraviolet crosslinking probably reflect the stable interaction with Rrp44 even when Rrp6 is present[13] (Extended Data Fig. 6a). Additional evidence for stochastic sampling of the two paths is evident by ultraviolet crosslinking under conditions of slight enzyme excess (Fig. 4d). As predicted based on the distributive and processive mechanisms of Rrp6 and Rrp44, respectively, crosslinked products are observed to Rrp44, Rrp6 and the S1/KH ring proteins at the earliest times, and this pattern is lost once most of the RNA finds its way to the Rrp44 active site.

Structural analysis of the Exo10[Rrp6] poly(A) complex suggests at least four potential paths past the S1/KH ring to Rrp6, although paths 1 and 2
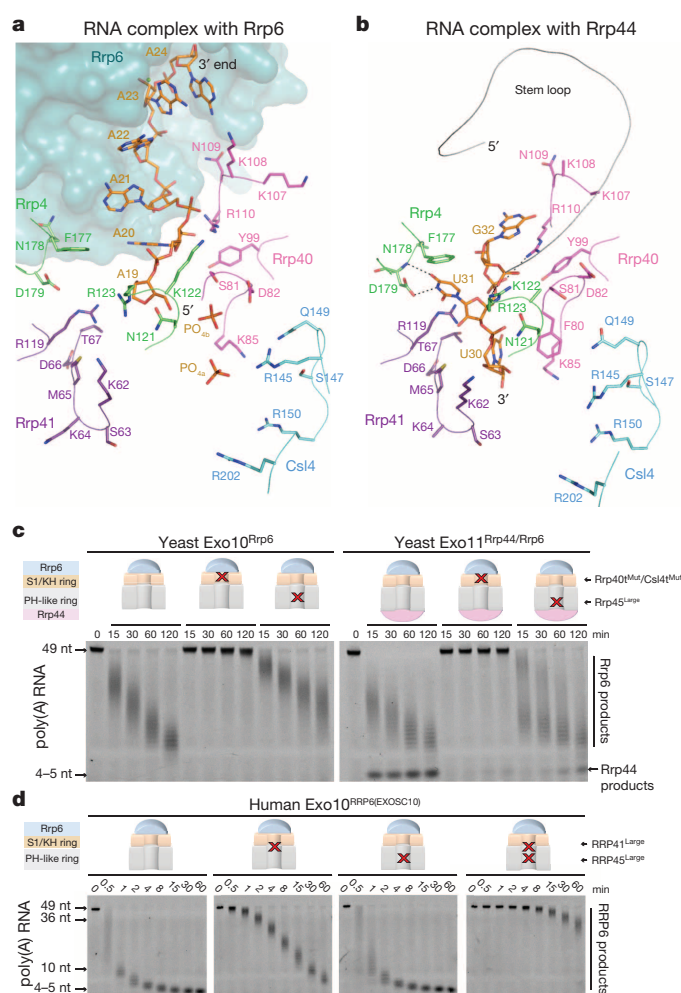
Figure 3 | Rrp6 activities are dependent on the S1/KH ring. a, b, Partially overlapping paths for RNA in complex with Exo10[Rrp6] (**a**) and Exo10[Rrp44 + Rrp6Cterm] (**b**) within the S1/KH ring. RNA and S1/KH side chains depicted as sticks, Rrp6 as a surface (**a**), the stem loop in Exo10[Rrp44 + Rrp6Cterm] as a ribbon (**b**). Colour-coded labels adjacent to subunits. **c,** RNA decay activities for yeast Exo10[Rrp6] and Exo11[Rrp44/Rrp6] with mutations in the S1/KH ring or a channel occluding insertion lower in the PH-like ring (Rrp45[Large]). **d,** RNA decay activities for human Exo10[RRP6] with insertions at the top (RRP41[Large]) or middle of the PH-like ring (RRP45[Large]). RNA decay assays were conducted at least three times for each of the samples with typical results depicted in **c** and **d**.
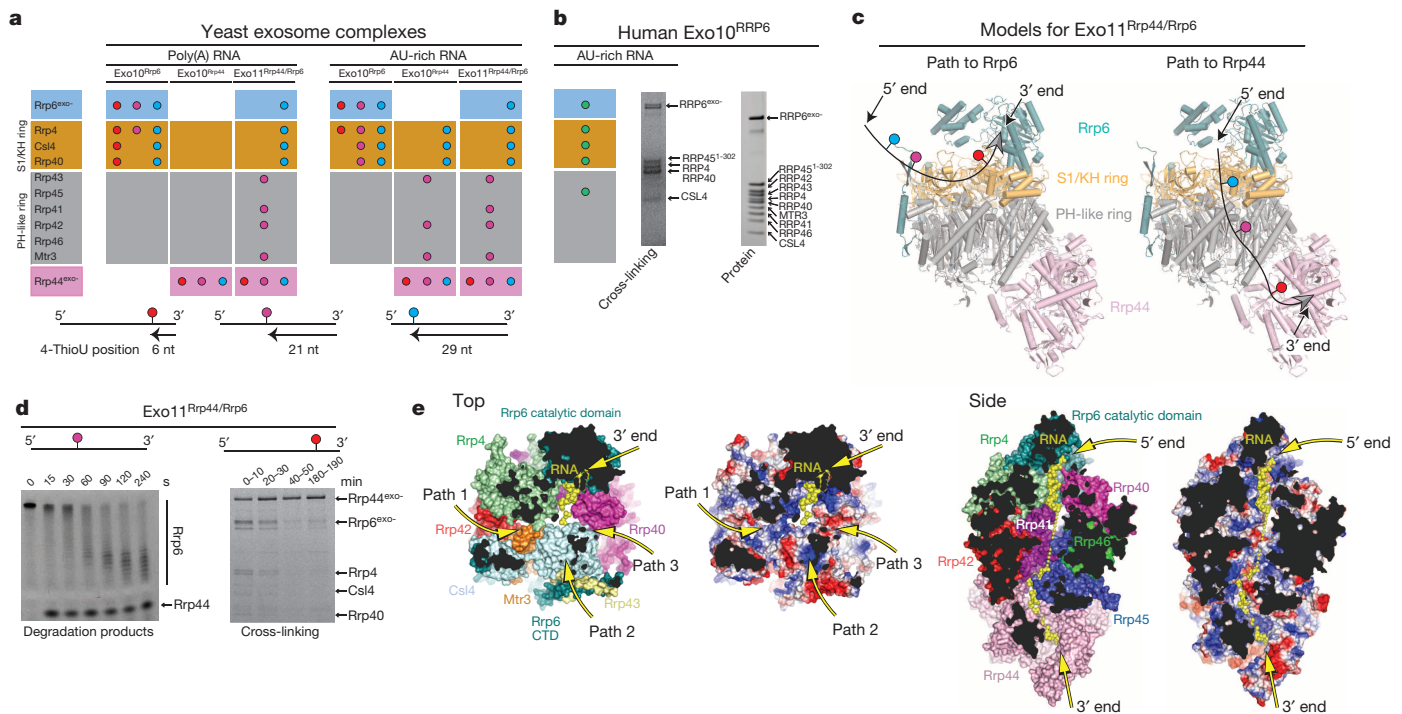
Figure 4 | RNA paths to Rrp6 and Rrp44. a, RNA crosslinking pattern schematic for yeast exosomes (Extended Data Fig. 6a). Subunits labelled, background coloured with Rrp6 (blue), S1/KH ring (orange), PH-like ring (grey) and Rrp44 (pink). 4-thioU positions indicated by coloured circles at 6, 21 or 29 nt from the 3′ end. **b**, RNA crosslinking pattern schematic for human Exo10^{RRP6} as in **a**. Crosslinks (left); proteins (right) (Extended Data Fig. 7). **c**, Exo11$^{Rrp44/Rrp6}$ cartoon. Subunits and rings coloured and labelled. RNA paths depicted by arrows, arrowhead at 3′ end with 4-thioU positions indicated as in **a**. **d**, RNA degradation at fivefold enzyme excess with Rrp6 and Rrp44 products (left) and time course for crosslinking to 4-thioU RNA 6 nt from

3′ end (right). **e**, Exo11$^{Rrp44/Rrp6}$ model indicating RNA paths to Rrp6 (left) or Rrp44 (right). Subunits coloured as in Fig. 1 next to surfaces coloured according to electrostatic potential[27]. Visible subunits and RNA (yellow) labelled in left panels. Three putative paths past the S1/KH ring to Rrp6 denoted by yellow arrows (Extended Data Fig. 8a). Side view shows an RNA path to Rrp44 (yellow arrow; 5′ and 3′ ends labelled) indicating that RNA could pass by Rrp6 (Extended Data Fig. 8b). RNA from PDB 4IFD was edited to remove the stem loop after nt 36. Assays in **a**, **b**, **d** conducted at least three times with typical results depicted.

seem most likely with respect to electrostatics and conservation (Fig. 4e; Extended Data Fig. 8a). Importantly, these paths are available in Exo11$^{Rrp44/Rrp6}$ as they do not involve surfaces from the PH-like ring or Rrp44. Modelling Rrp6 onto the Exo10^{Rrp44} RNA complex shows that the central channel is still accessible and that RNA paths to Rrp6 and Rrp44 are available in Exo11$^{Rrp44/Rrp6}$ (Fig. 4e; Extended Data Fig. 8b).

A wider channel in Exo10^{Rrp6}

Rrp6 can stimulate Rrp44 binding and decay activities[13], a phenomenon readily apparent in crosslinking to poly(A) RNA (Fig. 4a; Extended Data Fig. 6a). These results indicate that Rrp6 enhances RNA access to the PH-like ring and central channel. Structures of Exo10^{Rrp6} and Exo10$^{Rrp44+Rrp6Cterm}$ were compared to query differences that might account for this activity. Whereas the global architecture of the PH-like ring does not differ (Fig. 5d), the Exo9 channel widens in Exo10^{Rrp6} through movement of Rrp4, Rrp40 and Csl4 away from the central channel, widening the gap between Rrp4 and Rrp40 S1 domains by ~4 Å (Fig. 5a–c). Although Exo10^{Rrp6} and Exo10$^{Rrp44+Rrp6Cterm}$ structures are bound to RNA, the increase in channel width in Exo10^{Rrp6} might account for Rrp6-mediated stimulation of Rrp44, especially because Rrp44 and Rrp6 do not physically interact. It remains unclear how Rrp6 exerts this change or if an Exo10^{Rrp44} apo structure differs from its RNA-bound configuration, but it is clear that Rrp6 EXO and CTD are both required for this activity as addition of either element alone is not sufficient to stimulate Exo10^{Rrp44} when added *in trans* (Fig. 5e). These data suggest that the Rrp6 CTD is required to bring the catalytic module in proximity to the S1/KH ring, perhaps eliciting channel widening through EXO domain interaction with Rrp4 and Rrp40. Further data will be required to determine if channel widening is a regulated feature of Exo11$^{Rrp44/Rrp6}$ or if additional cytoplasmic factors elicit channel widening of Exo10^{Rrp44}.

Conclusions

The structure of Exo10^{Rrp6} shows Rrp6 positioned above the Exo9 S1/KH ring while the Exo10$^{Rrp44+Rrp6Cterm}$ structure shows Rrp44 below the PH-like ring. It is notable that Rrp6 activity is altered and becomes dependent on the S1/KH ring when associated with Exo9 (ref. 13). Although shorter RNAs may be directed to Rrp44 via a channel-independent "direct access" route[18], Rrp44 remains highly dependent on the integrity of the central channel throughout both S1/KH and PH-like rings. The dependency on the Exo9 core is remarkable given that both Rrp44 and Rrp6 active sites are exposed to solvent in Exo10^{Rrp6} and Exo10$^{Rrp44+Rrp6Cterm}$ complexes.

That a similar segment of the S1/KH ring is used to engage RNA in opposing directions suggests that overlapping paths to Rrp6 and Rrp44 may serve to commit the exosome to distributive or processive degradation depending on how a particular RNA substrate is delivered to the exosome. Because the paths overlap, the exosome would be unable to interact with another substrate until completing the task at hand. Whereas path choice seems stochastic *in vitro* (Fig. 4d), the nuclear cofactors Mpp6 (ref. 19), Rrp47 (ref. 8) and TRAMP[20] may bias selection of a particular path to facilitate transitions between editing, processing or degradation activities of the exosome.

Rrp6 can stimulate Rrp44 activities when associated with the Exo9 core, perhaps through widening of the S1/KH ring. Channel gating mechanisms as a point of regulation have been described in many systems including, but not limited to, the proteasome[21–25]. Whereas many degradation complexes place their catalytic activities in a protected compartment that is accessed by gating a substrate-responsive channel, the RNA exosome appears inside out, with its catalytic subunits located on the periphery of the exosome core. While the functional relevance of this architecture is not yet clear, it is evident that the exosome core can
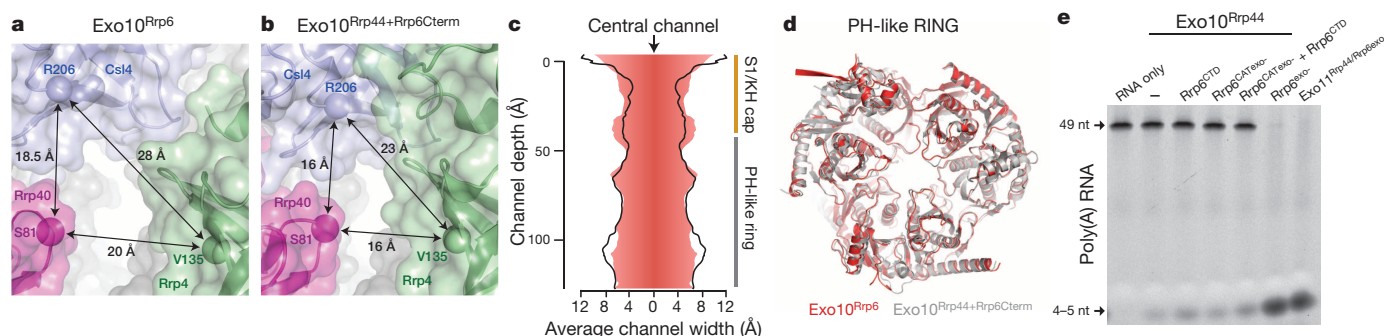
Figure 5 | S1/KH ring widening in Exo10^{Rrp6}. **a, b**, Close-up of the central channel looking down the S1/KH ring. Rrp4 (green), Rrp40 (magenta) and Csl4 (light blue) in cartoon with transparent surfaces from Exo10^{Rrp6} (**a**) and Exo10$^{Rrp44+Rrp6Cterm}$ (**b**). Distances indicated between respective labelled amino acids. **c**, Channel dimensions in Exo10^{Rrp6} (red) and Exo10$^{Rrp44+6Cterm}$ (black) calculated using HOLE[28]. **d**, Superposition of PH-like rings for Exo10^{Rrp6} (red) and Exo10$^{Rrp44+6Cterm}$ (grey). **e**, Rrp6^{exo-} (128–733), and not Rrp6^{CTD} (518–733) or Rrp6^{CAT} (128–518), is required to stimulate Rrp44 activity. RNA decay assays conducted at least three times with typical results depicted.

modulate the activities of its catalytic subunits by requiring RNA to pass through distinct elements of the central channel before being processed or degraded.

METHODS SUMMARY

RNA exosome subunits were expressed, purified and reconstituted as described[13,26]; see Methods for details. *S. cerevisiae* Exo10$^{Rrp6exo-}$ (12–14 mg ml^{-1}) was incubated with poly(A)$_{24}$ RNA (Invitrogen) at 1:1.1 molar ratio before crystallization. The structure was solved by molecular replacement. Mutant exosome subunits were generated by PCR-based site-directed mutagenesis. Wild-type and mutant exosome complexes were assayed for RNA degradation activities as described[13]. Ultraviolet crosslinking used catalytically-dead Rrp44 or Rrp6 using conditions described[13].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Wasmuth, E. V. & Lima, C. D. in *The Enzymes* (eds Chanfreau, G.F. & Tamanoi, F.) Ch. 31, 53–76 (Elsevier, 2012).
2. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. The exosome: A conserved eukaryotic RNA processing complex containing multiple 3′ to 5′ exoribonucleases. *Cell* **91,** 457–466 (1997).
3. Tomecki, R. *et al.* The human core exosome interacts with differentially localized processive RNases: hDis3 and hDis3L. *EMBO J.* **29,** 2342–2357 (2010).
4. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136,** 763–776 (2009).
5. Briggs, M. W., Burkard, K. T. & Butler, J. S. Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8S rRNA 3′ end formation. *J. Biol. Chem.* **273,** 13255–13263 (1998).
6. Yang, W., Lee, J. Y. & Nowotny, M. Making and breaking nucleic acids: two-Mg^{+2} ion catalysis and substrate specificity. *Mol. Cell* **22,** 5–13 (2006).
7. Beese, L. S. & Steitz, T. A. Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10,** 25–33 (1991).
8. Stead, J. A., Costello, J. L., Livingstone, M. J. & Mitchell, P. The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein. *Nucleic Acids Res.* **35,** 5556–5567 (2007).
9. Callahan, K. P. & Butler, J. S. Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p. *Nucleic Acids Res.* **36,** 6645–6655 (2008).
10. Midtgaard, S. F. *et al.* Structure of the nuclear exosome component Rrp6p reveals an interplay between the active site and the HRDC domain. *Proc. Natl Acad. Sci. USA* **103,** 11898–11903 (2006).
11. Januszyk, K., Liu, Q. & Lima, C. D. Activities of human RRP6 and structure of the human RRP6 catalytic domain. *RNA* **17,** 1566–1577 (2011).
12. Makino, D. L., Baumgärtner, M. & Conti, E. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* **495,** 70–75 (2013).
13. Wasmuth, E. V. & Lima, C. D. Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. *Mol. Cell* **48,** 133–144 (2012).
14. Assenholt, J. *et al.* Exonucleolysis is required for nuclear mRNA quality control in yeast THO mutants. *RNA* **14,** 2305–2313 (2008).
15. Liu, Q., Greimann, J. C. & Lima, C. D. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127,** 1223–1237 (2006).
16. Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E. & Conti, E. The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell* **139,** 547–559 (2009).
17. Drążkowska, K. *et al.* The RNA exosome complex central channel controls both exonuclease and endonuclease Dis3 activities *in vivo* and *in vitro. Nucleic Acids Res.* **41,** 3845–3858 (2013).
18. Liu, J. J. *et al.* Visualization of distinct substrate-recruitment pathways in the yeast exosome by EM. *Nature Struct. Mol. Biol.* **21,** 95–102 (2014).
19. Milligan, L. *et al.* A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. *Mol. Cell. Biol.* **28,** 5446–5457 (2008).
20. LaCava, J. *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121,** 713–724 (2005).
21. Makino, D. L., Halbach, F. & Conti, E. The RNA exosome and proteasome: common principles of degradation control. *Nature Rev. Mol. Cell Biol.* **14,** 654–660 (2013).
22. Kish-Trier, E. & Hill, C. P. Structural biology of the proteasome. *Annu. Rev. Biophys.* **42,** 29–49 (2013).
23. Sen, M. *et al.* The ClpXP protease unfolds substrates using a constant rate of pulling but different gears. *Cell* **155,** 636–646 (2013).
24. Effantin, G., Maurizi, M. R. & Steven, A. C. Binding of the ClpA unfoldase opens the axial gate of ClpP peptidase. *J. Biol. Chem.* **285,** 14834–14840 (2010).
25. Cha, S. S. *et al.* Crystal structure of Lon protease: molecular architecture of gated entry to a sequestered degradation chamber. *EMBO J.* **29,** 3520–3530 (2010).
26. Greimann, J. C. & Lima, C. D. Reconstitution of RNA exosomes from human and *Saccharomyces cerevisiae* cloning, expression, purification, and activity assays. *Methods Enzymol.* **448,** 185–210 (2008).
27. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.
28. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14,** 354–360 (1996).

Author Contributions E.V.W. and C.D.L. designed and E.V.W. performed experiments for the *S. cerevisiae* exosomes. E.V.W. and C.D.L. determined the structure. K.J. and C.D.L. designed and K.J. performed the experiments for the *Homo sapiens* exosome. E.V.W. and C.D.L. wrote the manuscript.

METHODS

Yeast exosome reconstitution and purification. Expression, purification, and reconstitution of recombinant exosome subunits and complexes have been previously described in detail[13,26]. In brief, the two catalytic subunits (Rrp6 and Rrp44) and nine non-catalytic subunits of the Exo9 core (Rrp41, Rrp45, Rrp42, Mtr3, Rrp46, Rrp43, Rrp4, Rrp40 and Csl4) were cloned as N-terminal $6\times$ histidine Smt3 fusions in pRSFDuet-1 expression vectors (Novagen), and expressed in the *Escherichia coli* expression strain BL21 (DE3) RIL (Stratagene) either as heterodimers (Smt3–Rrp41/Rrp45, Smt3–Rrp42/Mtr3, Smt3–Rrp46/Rrp43) or as distinct subunits (Smt3–Csl4, Smt3–Rrp4, Smt3–Rrp40, Smt3–Rrp6, Smt3–Rrp44). Generation of the Rrp40 and Csl4 point mutants was performed by PCR-mediated site-directed mutagenesis. For the mutant subunits, cells were grown in shaker flasks at 37 °C in Superbroth (Teknova), induced by cold shock on ice, addition of ethanol to a final concentration of 2% and 0.05 mM isopropyl-β-D-thiogalactoside (IPTG), and grown at 18 °C overnight. For all other proteins, cells were grown in shaker flasks at 37 °C in Superbroth and induced by cold shock on ice and addition of 0.4 mM IPTG, and then subjected to shaking at 18 °C overnight. Cells were collected and lysed as described previously. After high-speed centrifugation to pellet the insoluble material, the supernatant was loaded on a nickel-NTA column (Qiagen) and allowed to flow by gravity. After washing with 350 mM NaCl, 20 mM Tris pH 8.0, 1 mM BME, and 20 mM imidazole, the column was further washed with 350 mM NaCl, 50 mM KCl, 20 mM Tris pH 8.0, 1 mM BME, 2 mM ATP, 10 mM MgSO$_4$ to displace chaperone impurities. Protein was eluted with 350 mM NaCl, 20 mM Tris pH 8.0, 1 mM BME, and 250 mM imidazole and then purified by size exclusion chromatography. All subunits were purified on the Superdex 200 (GE) with the exception of Smt3–Csl4 and Smt3–Rrp40, which were purified on a Superdex 75 (GE). Wild-type and mutant subunits eluted at the same volumes on gel filtration. Only Smt3–Rrp6, Smt3–Rrp44, and Smt3–Rrp42/Mtr3 were subjected to overnight cleavage by the SUMO protease Ulp1 (ref. 29) with another step of purification on the Superdex 200 to remove the $6\times$ histidine Smt3 tags; Smt3–Rrp42/Mtr3 was cleaved by Ulp1 before formation of Exo9 as inclusion of this tag interferes with reconstitution. At this stage, subunits were concentrated to 6 to 12 mg ml^{-1} and stored at −80 °C. For reconstitution, Exo9 subunits were mixed together as Smt3 fusions, incubated on ice for 30 min, followed by addition of Ulp1 for another 30 min before overnight dialysis. After purification, exosomes were concentrated to 12–14 mg ml^{-1} and stored at −80 °C in a final buffer of 100 mM NaCl, 20 mM Tris pH 8.0, 0.1 mM MgCl$_2$, 1 mM TCEP.

Human exosome reconstitution and purification. For human, expression, purification and reconstitution strategies of the Exo9 core were identical except Smt3–RRP43/MTR3/RRP42 was expressed as a trimer, Smt3–RRP46 was expressed alone (as described previously[15,26]), and RRP45$^{1-302}$/RRP41 was purified with a non-cleavable N-terminal hexa-histidine tag on RRP45. RRP6 (EXOSC10, PM/SCL-100) residues 180 to 804 were PCR-amplified from a previously generated full-length expression construct[11], cloned into pET28a-Smt3 and transformed into BL21 (DE3) RIL. For expression, cells were grown in shaker flasks to ~0.6 OD$_{600\,nm}$, and induced overnight at 18 °C with 0.25 mM IPTG in the presence of 2% ethanol. Subsequent purification of human Smt3–RRP6$^{180-804}$ and reconstitution of HsExo10^{RRP6} is identical to that described for the yeast system. The purification method for the catalytic region, RRP6$^{180-606}$, used for core interaction studies has been described previously[11]. The channel occluding RRP41 and RRP45 mutants were engineered based on those described for *S. cerevisiae*[13], and were generated by PCR to encode electrostatic and steric loop insertions with the primary sequence GTGESEGESES between amino residues Gly 93 and Arg 94 of RRP45 and between Arg 62 and Ala63 of RRP41.

Crystallization and structure determination. Exosome samples were mixed with poly(A)$_{24}$ RNA (Invitrogen) in a 1:1.1 molar ratio and incubated on ice for 1 h before crystallization. Crystals grew at 18 °C by vapour diffusion in either sitting (Greiner Bio One, Crystalquick) or hanging (Hampton Research, VDX) drop formats, in 7–11% PEG3350, 100 mM MES pH 6.7, 4–15% MPD, typically taking 3 to 5 days to appear. Crystals were harvested within 2 weeks. Partial degradation of RNA (20–24 nt) was observed in washed crystals (Extended Data Fig. 1d). For cryoprotection, well solution was replaced with crystallization buffer augmented with 25–30% MPD and crystals were incubated for three days before harvesting; one day before harvest, trays were transferred to 4 °C. A light polarizer was used to exclude multiple crystals and to identify crystals with single regions, which were broken off and flash frozen in liquid nitrogen for data collection. X-ray diffraction data were collected at the Advanced Photon Source 24-ID-E and 24-ID-C beam lines, and the National Synchrotron Light Source X29 beam line. Data was obtained from a single crystal diffracted at NSLS X29 at a wavelength of 1.075 Å at 100 K. Data were processed using HKL2000[30] and statistics reported in Extended Data Table 1 were obtained using Phenix[31] including CC$_{1/2}$ (%) and CC* (%) values of 99.9 (33.8) and 100.0 (71.1) for data between 50–3.3 (3.42–3.3) Å. The structure was solved by molecular replacement using Phaser[32]

and coordinates of human apo Exo9 (PDB: 2NN6) and yeast Rrp6^{CAT} (PDB: 2HBL) as search models followed by docking yeast Exo10$^{Rrp44+6Cterm}$ (PDB: 4IFD) (Extended Data Table 1). The structure is refined to R/R_{free} values of 0.227/0.265. The final model includes six nucleotides of the poly(A) 24mer, and two phosphate ions that likely represent the RNA backbone (Extended Data Fig. 1a–c), and 2,751 of the 3,156 amino acids present in the crystal (Extended Data Table 2). The asymmetric unit contains one complex. Iterative rounds of refinement were accomplished using Phenix[31]. RNA and side chains were manually built using O[33] and Coot[34]. The model was initially refined using secondary structure restraints in conjunction with positional refinement followed by individual B-factor refinement. Secondary structure restraints were released and a final round of positional and B-factor refinement was performed followed by refinement of TLS parameters that resulted in a further decrease in R and R_{free} values. Figures depicting the structure were prepared with Pymol[27]. Surface conservation was calculated using ConSurf[35]. Structure quality was assessed using MolProbity[36] indicating the model has excellent geometry with 95.7% in favoured and 100% in allowed regions of Ramachandran space. The structure also scored in the 100th percentile for the Clash and MolProbity scores.

RNA biochemistry. For yeast exosomes, unless otherwise noted, exoribonuclease assays were performed under multiple turnover conditions as described previously[13]. In brief, 10 nM synthetic 49-nt RNAs (Invitrogen) bearing 5′ fluorescein labels were incubated at 30 °C with 1 nM RNA exosome for various time points, and RNA degradation was monitored by resolving reaction intermediates by denaturing TBE-PAGE (Invitrogen), and detected using a Fuji FLA-5000 scanner (FITC filter). A similar protocol was followed for human exosome complexes, the exception being that these assays were performed at 37 °C, with 5 nM of the 49-nt RNA and 50 nM HsExo10^{RRP6}. For mix-in experiments (Fig. 5e), twofold molar excess of Rrp6 protein was incubated with Exo10^{Rrp44} on ice for 1 h before initiating RNA decay by addition of RNA. For steady state ultraviolet-RNA crosslinking to 4-thioU-bearing RNAs, 500 nM of exosome complexes with mutations in the exoribonuclease active sites (D238N for Rrp6, D551N for Rrp44) were incubated for 20 min on ice with 50 nM 36-nt 5′ fluorescein RNAs with a single internal 4-thioU (Thermo Scientific) in a 70 µl reaction volume. Binding buffer includes 50 mM KCl, 20 mM Tris pH 8.0, 10 mM DTT, 0.5 mM MgCl$_2$. Crosslinking was performed by subjecting the RNA-exosome mixture to long-range ultraviolet (365 nm) for 15 min in the dark using a 4 W handheld lamp. For the time course described in Fig. 4d, 350 nM of exosome was incubated with 150 nM of RNA, and the binding reactions allowed to proceed for indicated times before initiating crosslinking for 10 min. 15 µl was quenched with LDS loading buffer, and the crosslinked products were separated by SDS–PAGE (Invitrogen) and visualized with a Fuji FLA-5000 scanner (FITC filter). 4-thioU RNA included the following sequences, each of which contained a 5′-fluorescein: 5′-Fl-AAUUAU4thioUUAUUAUUUAUUUAUUAUUUAUUUAUUUAA; 5′-Fl-AAUUAUUUAUUAUU4thioUAUUUAUUAUUUAUUUAUUUAA; 5′-Fl-AAUUAUUUAUUAUUUAUUUAUUAUUUAUU4thioUAUUUAA; 5′-Fl-AAAAAA4thioUAAAAAAAAAAAAAAAAAAAAAAAAAAAAA; 5′-Fl-AAAAAAAAAAAAAA4thioUAAAAAAAAAAAAAAAAAAAAAA; 5′-Fl-AAAAAAAAAAAAAAAAAAAAAAAAAA4thioUAAAAAA. For ultraviolet crosslinking to human exosome complexes, samples were incubated with 250 nM of the AU-rich RNA and protein in binding buffer for 60 min on ice, and then placed in a ultraviolet-Stratalinker (Stragene) and subjected to 300,000 µJ of short-wave (254 nm) ultraviolet radiation. Samples were quenched with LDS loading buffer, and crosslinked products were separated and visualized as described above.

29. Mossessova, E. & Lima, C. D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **5,** 865–876 (2000).
30. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* (eds Carter, C. W. Jr & Sweet, R. M.) Vol. 276, 307–326 (Academic Press, 1997).
31. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
33. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
35. Ashkenazy, H., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** W529–W533 (2010).
36. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
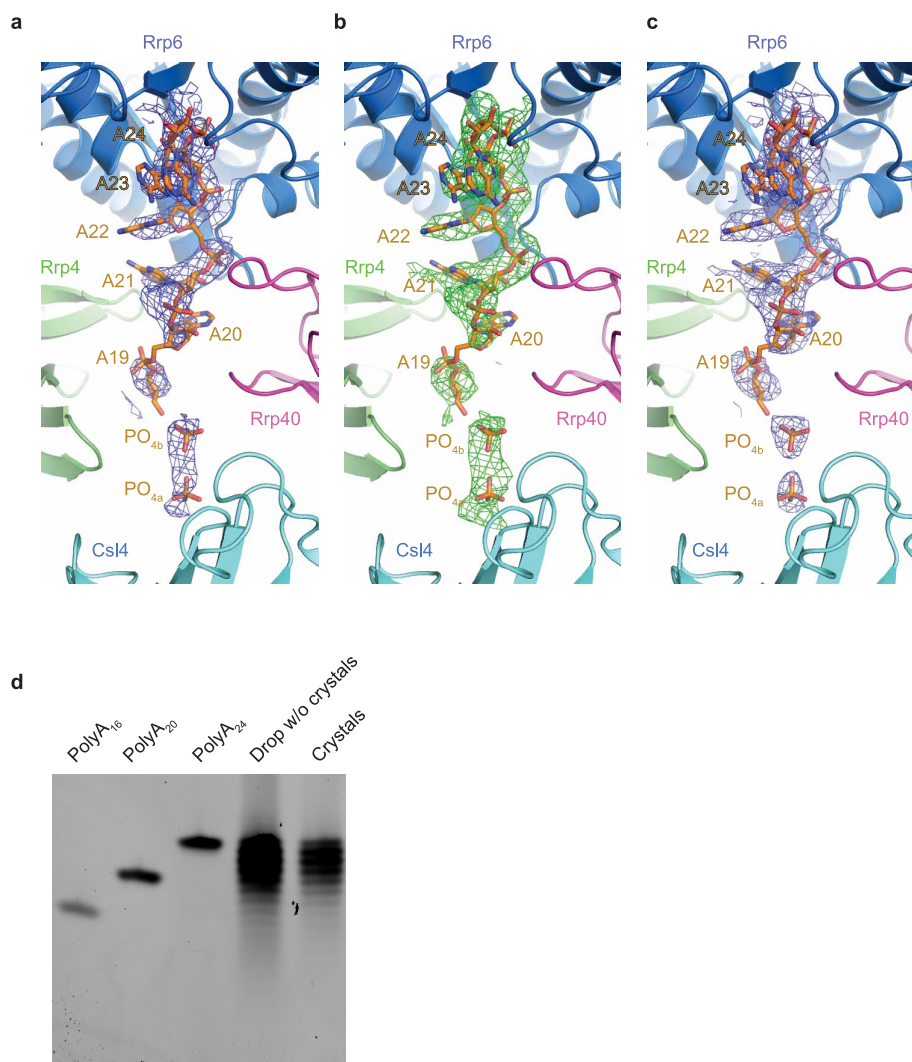
Extended Data Figure 1 | View highlighting electron densities covering RNA in the Exo10^{Rrp6}–poly(A) structure. a, Simulated annealing $2F_o - F_c$ omit map contoured at 1.0σ. **b**, Simulated annealing $F_o - F_c$ omit map contoured at 2.0σ. **c**, Final $2F_o - F_c$ omit map contoured at 1.0σ. RNA and phosphate oxygen, nitrogen, carbon and phosphate atoms are coloured red, blue, orange and dark orange, respectively. Electron density maps shown as wire mesh coloured blue for $2F_o - F_c$ maps and green for the $F_o - F_c$ map.

Rrp6, Rrp40, Rrp4 and Csl4 are labelled and shown in cartoon representation in teal, magenta, green and light blue, respectively. **d**, 20 to 24 nucleotides of poly(A) RNA are present in crystals of Exo10^{Rrp6}. Crystals of Exo10^{Rrp6} bound to poly(A)$_{24}$ RNA were first washed by two rounds of transfer to 1 µl well solution, then dissolved in water and TBE-urea sample buffer, and analysed by 15% TBE-urea PAGE. A drop without crystals was run as a control. RNA was stained by Sybr Gold.
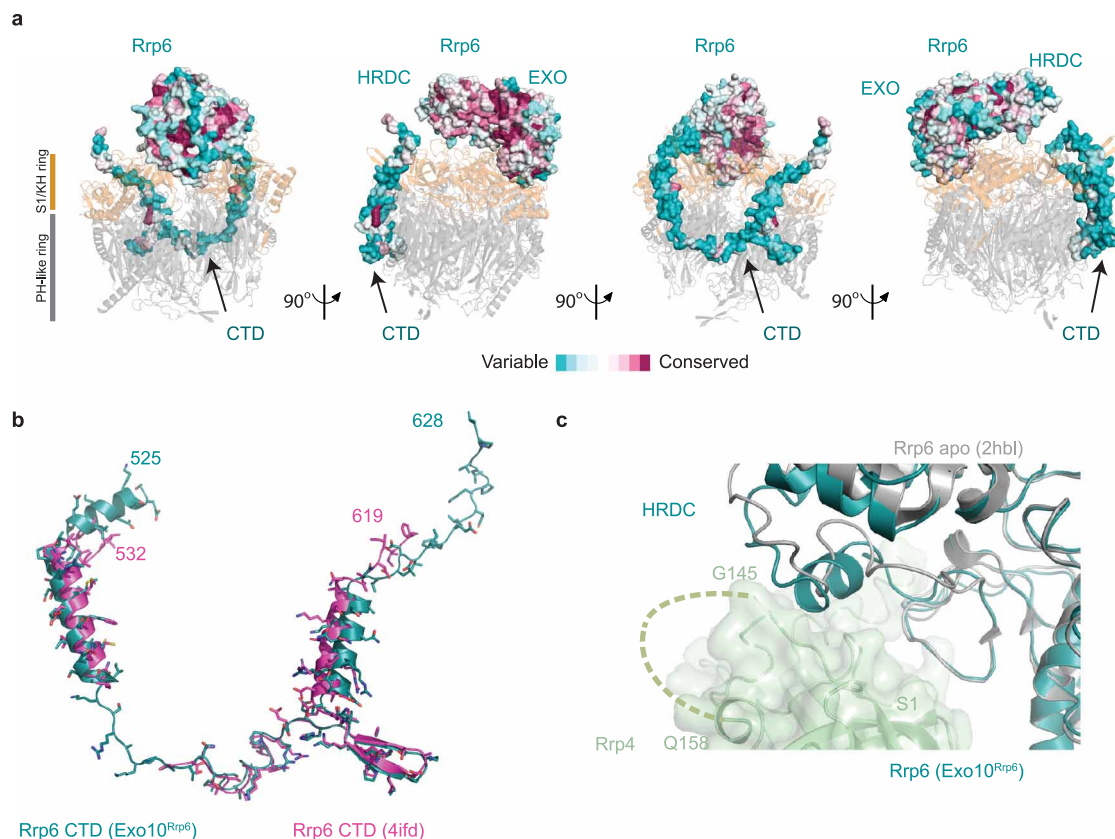
Extended Data Figure 2 | Rrp6 features within Exo10^{Rrp6}. a, The primary sequence of the catalytic module of Rrp6 (EXO and HRDC) is highly conserved, whereas sequence conservation within the CTD varies. Rrp6 in Exo10^{Rrp6} is represented as a surface, and coloured according to sequence conservation as calculated by ConSurf[35], coloured from red (highly conserved) to blue (variable). The Exo9 core is depicted as a transparent cartoon, with the S1/KH ring in orange, and the PH-like ring in grey. **b,** Superposed structures of the Rrp6 CTD from Exo10^{Rrp6} (residues 525–628; teal) and Exo10$^{Rrp44 + Rrp6Cterm}$ (residues 532–557, 565–619; magenta; PDB 4IFD) reveal overall similarity for residues within the respective models. **c,** The conserved linker between the EXO and HRDC domain of Rrp6 adopts an α-helix when associated with the exosome core. Superposition of the Rrp6 catalytic domains from the structures of Exo10^{Rrp6} and Rrp6–AMP (PDB: 2HBL) shows residues Pro 424 to Asn 433 rearrange from a loop in the AMP-bound structure to a helix that is adjacent to a conserved, albeit disordered, loop in the S1 domain of Rrp4. These residues were previously thought to comprise a protein interaction module[10]. Rrp6 is represented as cartoons, with Rrp6 from the Exo10^{Rrp6} structure in teal, and the Rrp6–AMP structure in grey. Rrp4 is in light green and depicted as a cartoon with a transparent surface, with the S1 loop representing residues Gly 145 to Gln 158 shown as a dashed line.

a

Rrp6 domains

CAT

CTD

PMC2NT EXO HRDC

ScRrp6$^{518-733}$
ScRrp6$^{128-518}$
ScRrp6$^{128-733}$
HsRRP6$^{180-606}$
HsRRP6$^{180-804}$

b

UV$_{280}$ absorbance

hExo10^6

hRrp6

Elution volume

Subunit (kDa)

RRP6$^{180-804}$ (72)

RRP45 (34)
RRP42 (32)
RRP43 (30)
RRP40 (30) / hRrp4 (33)
MTR3 (28)
RRP41 (26)
RRP46 (25)
CSL4 (23)

Coomassie stained SDS-PAGE

c

RRP6$^{180-606}$
HsExo9
HsExo9 + RRP6$^{180-606}$
HsExo9 + RRP6$^{180-804}$

UV$_{280}$ absorbance

12.5 15.0 17.5 (mL)

RRP6$^{180-606}$: ← RRP6$^{180-606}$

HsExo9: RRP45, RRP42, RRP43, RRP40/RRP4, MTR3, RRP41, RRP46, CSL4

HsExo9 + RRP6$^{180-606}$: ← RRP6$^{180-606}$ / RRP45, RRP42, RRP43, RRP40/RRP4, MTR3, RRP41, RRP46, CSL4

HsExo9 + RRP6$^{180-804}$: ← RRP6$^{180-804}$ / RRP45, RRP42, RRP43, RRP40/RRP4, MTR3, RRP41, RRP46, CSL4, ← Smt3

d

ScExo10^{44}
ScExo10^{44} + Rrp6$^{128-518}$
ScExo10^{44} + Rrp6$^{518-733}$
ScExo10^{44} + Rrp6$^{128-518}$ + Rrp6$^{518-733}$
ScExo10^{44} + Rrp6$^{128-733}$

UV$_{280}$ absorbance

12.5 15.0 17.5 (mL)

Load

ScExo10^{44} + Rrp6$^{128-518}$: ←Rrp44 / Rrp6$^{128-518}$ / ←Rrp43/Rrp6$^{128-518}$, ←Rrp4, ←Rrp45, ←Csl4, ←Rrp41/42/46, ←Mtr3/Rrp40

ScExo10^{44} + Rrp6$^{518-733}$: ←Rrp44 / ←Rrp43, ←Rrp4, ←Rrp45, ←Csl4, ←Rrp41/42/46, ←Mtr3/Rrp40, ←Rrp6$^{518-733}$ / Rrp6$^{518-733}$ / Rrp6$^{518-733}$ / ←Smt3

ScExo10^{44} + Rrp6$^{128-518}$ + Rrp6$^{518-733}$: ←Rrp44 / Rrp6$^{128-518}$ / ←Rrp43/Rrp6$^{128-518}$, ←Rrp4, ←Rrp45, ←Csl4, ←Rrp41/42/46, ←Mtr3/Rrp40, ←Rrp6$^{518-733}$, Rrp6$^{518-733}$ / Rrp6$^{518-733}$ / ←Smt3

ScExo10^{44} + Rrp6$^{128-733}$: ←Rrp44 / Rrp6$^{128-733}$ / ←Rrp6$^{128-733}$ / ←Rrp43, ←Rrp4, ←Rrp45, ←Csl4, ←Rrp41/42/46, ←Mtr3/Rrp40

Extended Data Figure 3 | The Rrp6 CTD is required for interaction with human and yeast RNA exosome cores. **a**, Schematic representation of Rrp6 including from N to C terminus the PMC2NT, EXO, HRDC and C-terminal (CTD) domains. EXO and HRDC comprise Rrp6^{CAT}. Below are lines representing individual elements used in gel-filtration analysis in **c** and **d** along with amino acid numbering specific for *S. cerevisiae* (Sc) Rrp6 and *Homo sapiens* (Hs) RRP6. **b**, Gel-filtration profile for reconstituted human Exo10^{RRP6} (left) with SDS–PAGE analysis of the peak fractions (right). Subunits labelled in capital letters corresponding to yeast nomenclature for clarity. Apparent molecular weights denoted in kilodaltons. **c**, Interaction studies of human RRP6 with the HsExo9. The human RRP6 catalytic domain (residues 180–606) does not interact well with human Exo9; however, addition of RRP6 180–804 suggests that CTD residues 607–804 are important for stable interaction with Exo9. Reconstitution experiments were performed as described above on mixtures of HsExo9, HsExo9 with twofold excess of RRP6$^{180–606}$ (the catalytic region), HsExo9 with twofold excess of RRP6$^{180–804}$ (the catalytic region

with a portion of the carboxy terminus), and RRP6$^{180–606}$. Gel filtration analysis on RRP6$^{180–804}$ in isolation was not possible as this construct is unstable in the absence of the SMT3 tag. Mixtures (0.15 ml at ~2 mg ml^{-1}) were dialysed overnight at 4 °C against reconstitution buffer (50 mM NaCl, 20 mM Tris-HCl pH 8.0, and 10 mM DTT). Mixtures were applied to a Superose 6 10/300 GL column (G.E. Health Sciences) equilibrated with reconstitution buffer with the resulting ultraviolet traces (upper panels) and analysed fractions (lower panel). **d**, Interaction studies of budding yeast Rrp6 with ScExo10^{Rrp44}. Rrp6 (residues 128–733) and Rrp6^{CTD}, but not Rrp6^{CAT}, interact with Exo10^{Rrp44}. Rrp6 (residues 128–733), Rrp6^{CAT} (residues 128–518), the Rrp6^{CTD} (residues 518–733), or both were incubated with Exo10^{Rrp44} for one hour on ice before analysis by size exclusion chromatography (Superose 6) in reconstitution buffer. Rrp6 and Rrp6^{CAT} were added in 1.5-fold molar excess to Exo10^{Rrp44}, while Rrp6^{CTD} was added in threefold molar excess. For panel **c** and **d**, fractions were analysed by SDS–PAGE and protein detected with Sypro Ruby.

a *S. cerevisiae* Rrp6 with RNA

b *H. sapiens* RRP6 with RNA (model)

c yRrp6cat AMP complex

d Hybrid model for Rrp6 active site
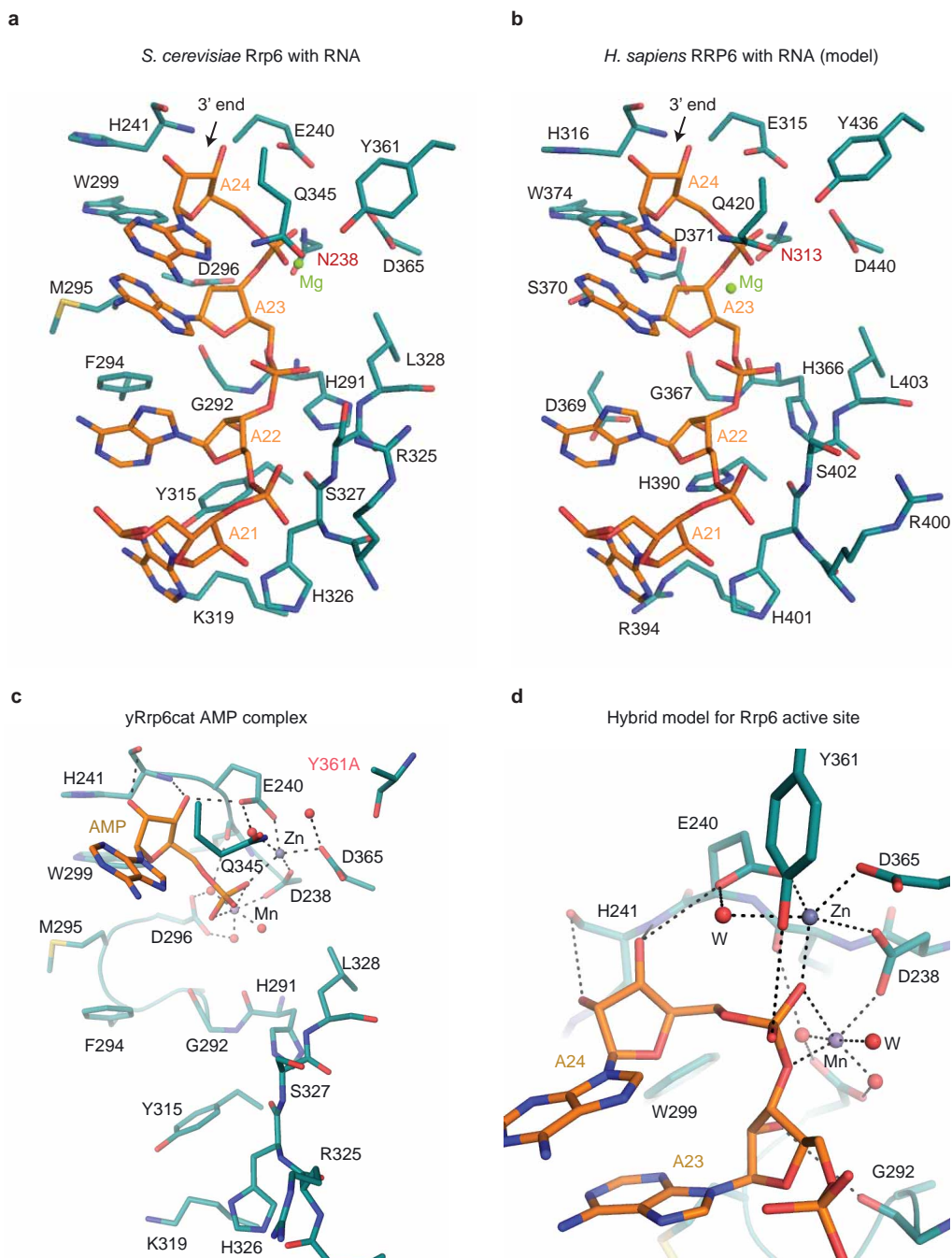
Extended Data Figure 4 | RNA interactions within the Rrp6 active site.
a, **b**, View of RNA nucleotides 24 to 20 from the 3′ end in the active site of
S. cerevisiae Rrp6 from the structure of Exo10^{Rrp6} (**a**) and the same RNA
modelled into the active site of *H. sapiens* RRP6 (PDB: 3SAF) based on
superposing the respective EXO domains (**b**). Residues involved in RNA
contacts in the structure of yeast Exo10^{Rrp6} are shown as teal sticks in **a**; the
corresponding residues in human are shown in **b**. In both structures, a catalytic
aspartate is mutated to asparagine and labelled in red (N238 in yeast; N313
in human). RNA is shown as orange sticks, and magnesium ions are

represented as green spheres. **c**, Structure of the yeast Rrp6 catalytic domain
in complex with zinc (Zn; blue sphere) and manganese (Mn; yellow sphere) in
complex with AMP (PDB 2HBL). Active site residues are depicted as in
a, **b**. The conserved Tyr 361 side chain was mutated to alanine in this structure
(labelled Y361A in red). **d**, Model of the Rrp6 active site constructed by
superposing the 3′ nucleotide in the Exo10^{Rrp6} poly(A) structure to the AMP
ligand from PDB 2HBL. Amino acid side chains are labelled and putative
interactions between metal ligands, active site residues and RNA are depicted
by dashed lines.
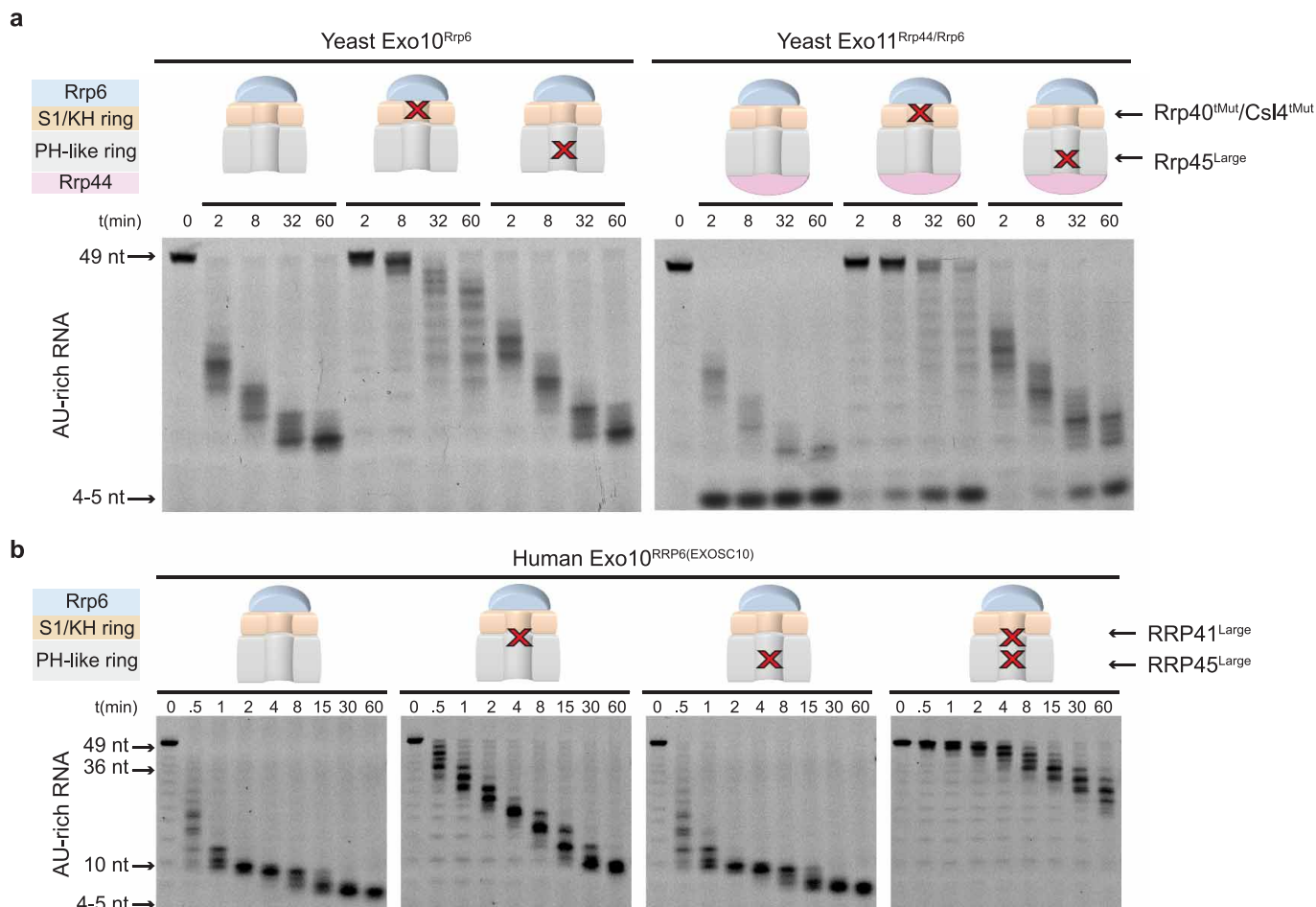
Extended Data Figure 5 | Central channel mutations and their impact on Rrp6 degradation of AU-rich RNA. **a**, 1 nM of reconstituted wild-type and mutant yeast Exo10^{Rrp6} and Exo11$^{Rrp44/Rrp6}$ were incubated with 10 nM 49-nt 5′ fluorescein AU-rich RNA at 30 °C. Reaction intermediates were analysed after indicated time points by denaturing PAGE and imaged with a fluoroimager. Rrp6 activity decreases in the presence of the Rrp40 K107E/K108E/R110D and Csl4 R145A/R150A/R202D mutations in both Exo10^{Rrp6} and Exo11$^{Rrp44/Rrp6}$ but is mostly unaffected by the channel occlusion in the Rrp45^{Large} exosome. Rrp44 exoribonuclease activity is attenuated by mutations in both the S1/KH and PH-like ring in Exo11$^{Rrp44/Rrp6}$. **b**, Exoribonuclease activities of wild-type and mutant forms of human Exo10^{RRP6} were performed with 5 nM 49-nt 5′ fluorescein AU-rich RNA and 50 nM protein at 37 °C. Reaction intermediates were analysed after indicated time points as described in **a**. Mutation studies reveal that the human RRP6 exoribonucleolytic activity decreases in the presence of a channel occlusion near the S1/KH ring (RRP41^{Large}), while a channel occlusion deeper in the central channel (RRP45^{Large}) of HsExo10^{RRP6} has little effect on activity.
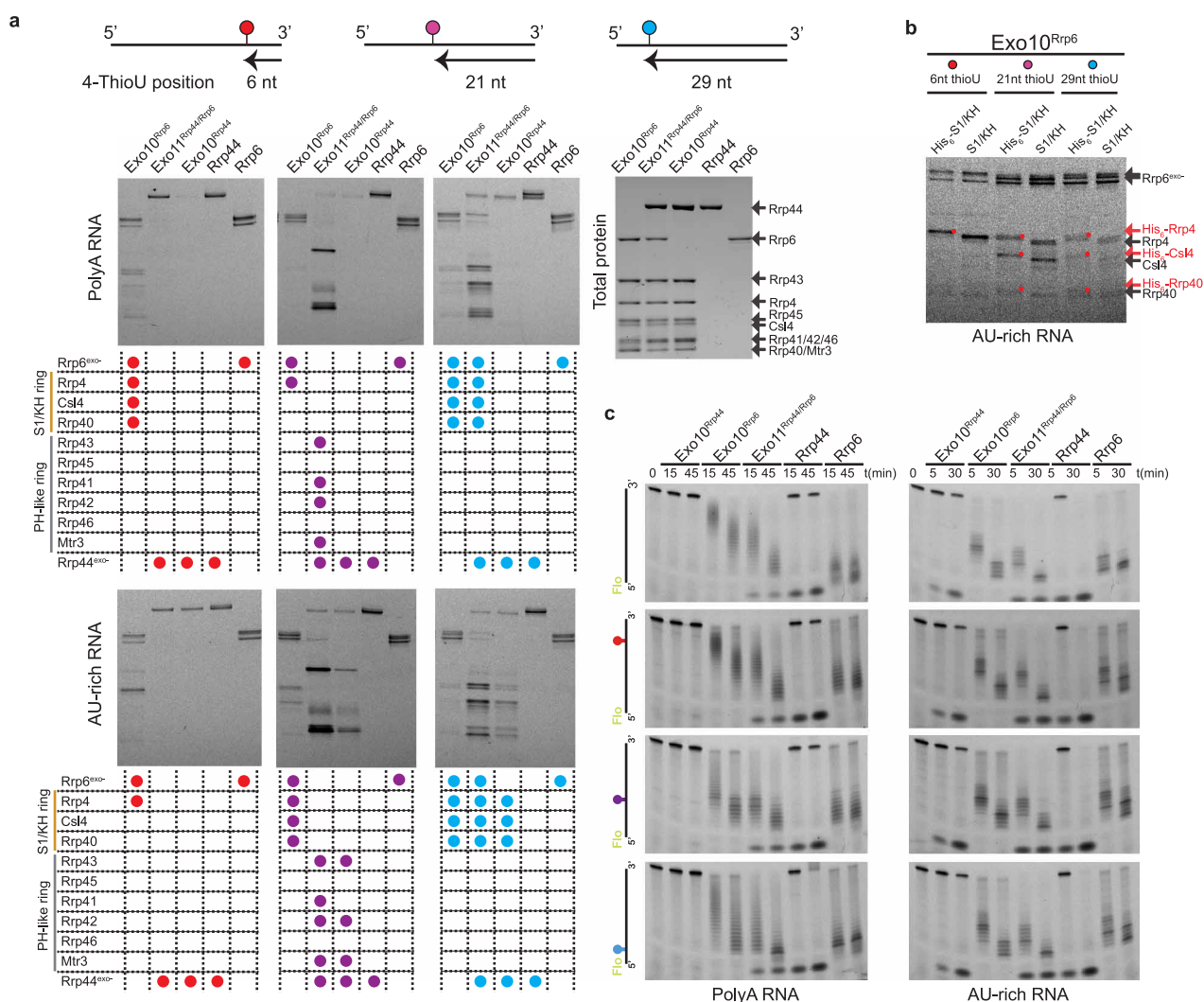
Extended Data Figure 6 | Ultraviolet crosslinking with 36-nt 4-thioU RNAs for yeast exosome complexes. **a**, Raw data used to generate Fig. 4a in main text for poly(A) and AU-rich RNA. See Methods for crosslinking details. Presence or absence of crosslinks to given subunits is summarized in the table and denoted by colour-coded dots, with red corresponding to 4-thioU 6 nt from the 3′ end; purple for 4-thioU 21 nt from the 3′ end; and blue for 4-thioU 29 nt from the 3′ end. **b**, RNA–protein adducts observed in ultraviolet crosslinking to Exo10^{Rrp6} correspond to the S1/KH cap proteins. Exo10^{Rrp6} bearing N-terminal hexahistine tags on the S1/KH cap proteins Rrp4, Rrp40 and Csl4 (ref. 13) or tag-free variants were crosslinked to the three 4-thioU

RNAs described above, and resolved by SDS–PAGE and imaged with a fluoroimager. In comparing the His$_6$-tagged and tagless complexes, mobility shifts are apparent in crosslinks corresponding to the S1/KH cap proteins. **c**, 4-thioU RNAs are bona fide decay substrates for exosome complexes and catalytic subunits, Rrp6 and Rrp44. Conditions using substrate excess (10 nM RNA, 1 nM enzyme), 4-thioU poly(A) RNAs (left) and AU-rich RNAs (right) are comparably degraded to corresponding 36-nt RNAs lacking 4-thioU. Reactions were performed at 30 °C and stopped after indicated time points and resolved by denaturing PAGE and imaged with a fluoroimager. 4-thioU RNAs are colour-coded as in **a** and **b**.
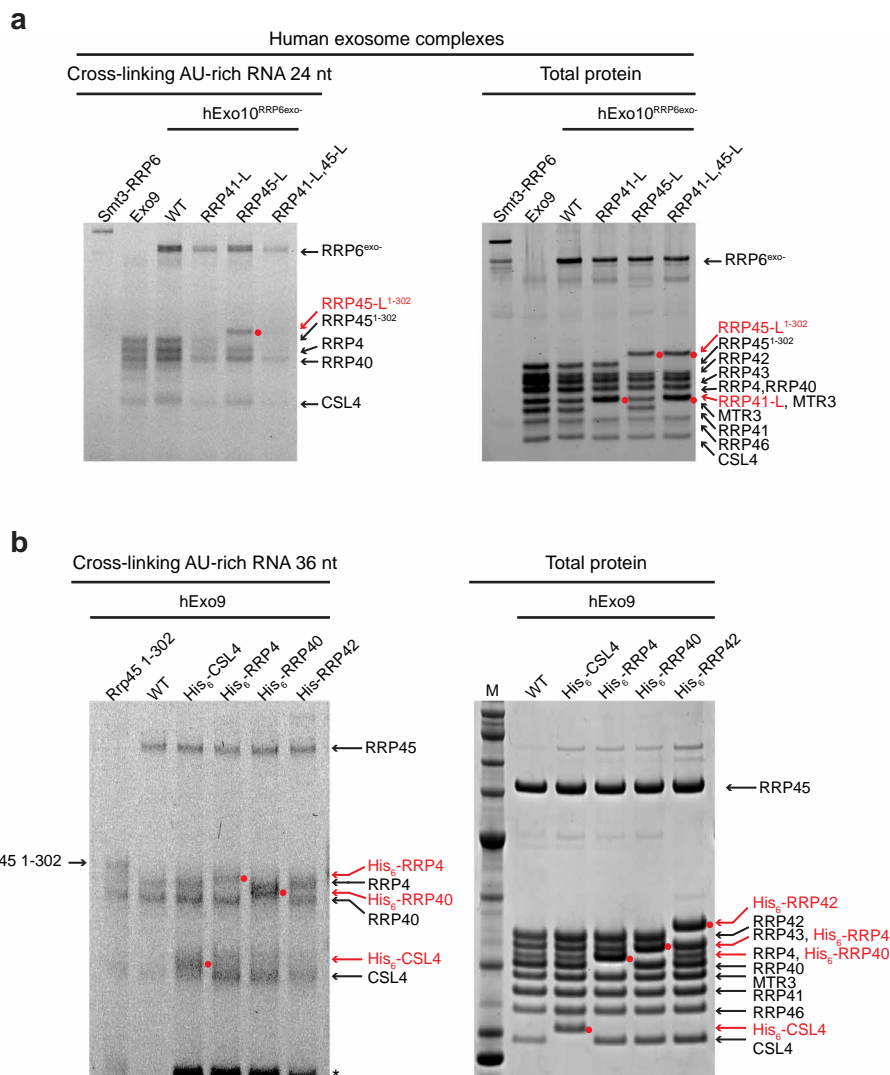
Extended Data Figure 7 | Ultraviolet crosslinking to AU-rich RNA to components of human Exo10^{RRP6}. a, Ultraviolet crosslinking of RNA to HsExo10^{RRP6} with open and occluded channels. Human RRP6 (Smt3–RRP6$^{180-804}$), HsExo9 and wild type as well as three different pore occlusion mutants of HsExo10^{RRP6} were incubated and ultraviolet crosslinked to a 24-nt fluorescein-labelled AU-rich RNA substrate. Samples were analysed for the presence of RNA-protein crosslinks (left panel) or integrity of protein via Sypro staining (right panel). Mutant forms of HsExo10^{RRP6} with insertion loops in RRP41-L and RRP45-L are labelled with red asterisks. b, Identification of ultraviolet-induced crosslinks. Identification of RNA-protein crosslinks

between RNA to cap components of the human Exo9 core (RRP4, RRP40, and CSL4) was facilitated by reconstituting His$_6$-tagged versions of these subunits into HsExo9 complexes, performing crosslinking assays with a 36-nt fluorescein-labelled AU-rich RNA, and detecting slower mobility for the His$_6$-tagged crosslinked adducts. Similarly, identification of RRP45-RNA adducts was based on changes in adduct mobility when using isoforms with wild-type carboxy termini (RRP45$^{1-456}$) and the long insertions (RRP45-L$^{1-302}$). RRP6–RNA adducts were determined by crosslinking of a Smt3–RRP6$^{180-804}$ fusion and comparison of 9- and 10-subunit exosomes (in a).
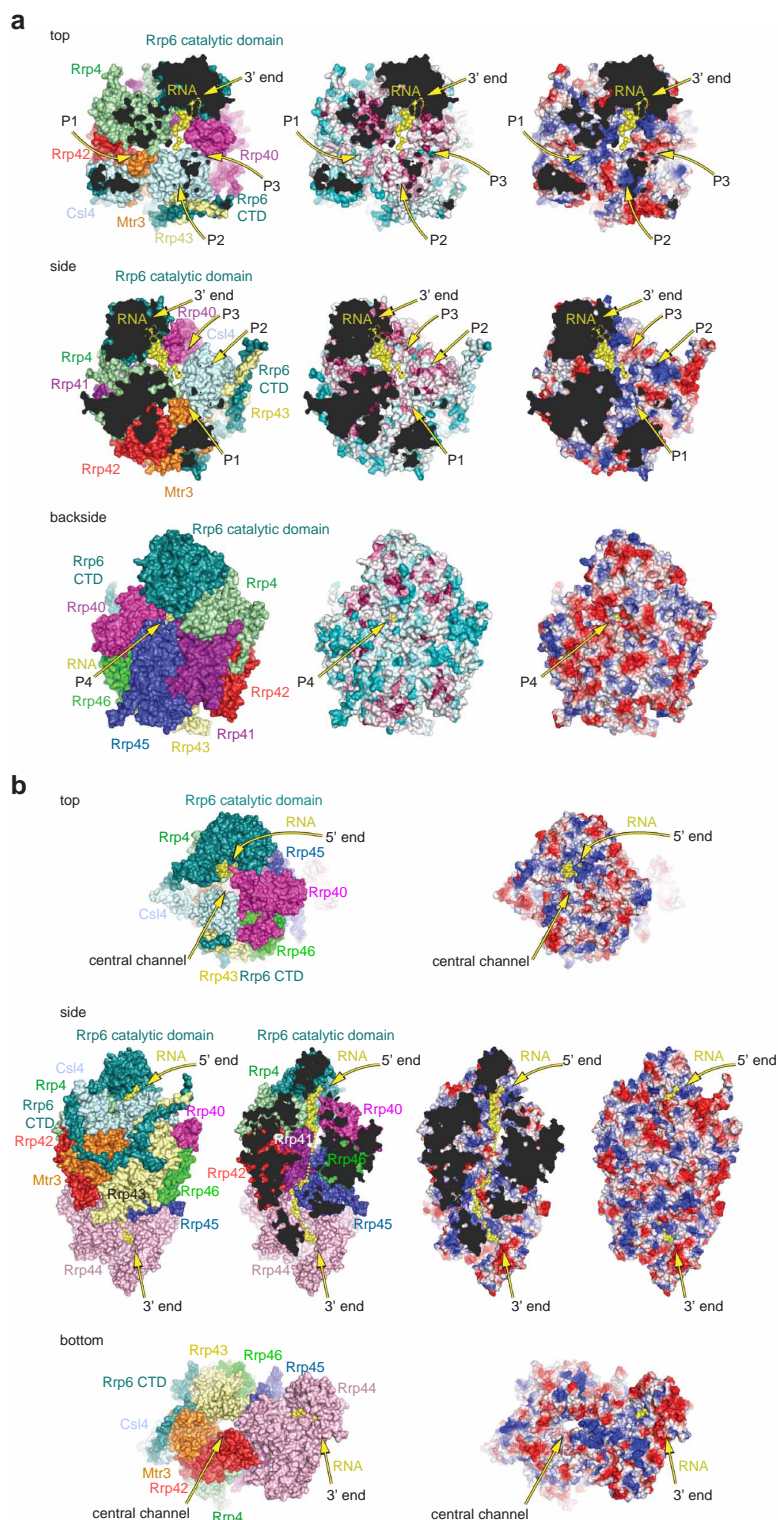
Extended Data Figure 8 | RNA paths to Rrp6 and Rrp44. Structure of an Exo11$^{Rrp44/Rrp6}$ model derived by superposing the PH-like ring subunits in Exo10^{Rrp6} and Exo10$^{Rrp44+6Cterm}$ (PDB 4IFD). **a**, RNA paths to the Rrp6 active site. RNA (yellow) shown as derived from the Exo10^{Rrp6} poly(A) complex. Three views are provided from the top, side and backside of the complex with subunits coloured as in Fig. 1 in the main text (left), surfaces coloured according to conservation as calculated by ConSurf[35] (middle) from red (highly conserved) to blue (variable), and surfaces coloured according to electrostatic potential as calculated by Pymol[27] (right). Subunits that are visible are labelled in the left panel as is the RNA. Three putative paths that traverse the S1/KH ring into the Rrp6 active site are denoted by yellow arrows and labels (P1, P2, P3) in the top and side views and a fourth path (P4) under the Rrp6 catalytic domain is denoted in the backside view. Based on conservation, electrostatics and crosslinking we deem P1 or P2 as the most likely path of ingress for the incoming RNA substrate. The top and side views use cutaway surfaces to illustrate the putative paths and surface properties. **b**, RNA paths to the Rrp44 active site. Subunits and surfaces depicted as in **a** with the exception that the RNA is now derived from the structure of Exo10$^{Rrp44+6Cterm}$ (PDB 4IFD) with the RNA edited to remove the stem loop after nucleotide 36. Top, side and bottom views are depicted with the RNA path indicated by a yellow arrow with 5′ and 3′ ends labelled. The position of the central channel is also labelled and indicated by a yellow arrow in the top and bottom views. The top and side views show that RNA could pass by Rrp6 to penetrate the S1/KH and PH-like ring central channel.

Extended Data Table 1 | Data collection and refinement statistics

	Exo10^{Rrp6}/polyA
Data collection	
Space group	P2$_1$2$_1$2
Cell dimensions	
a, b, c (Å)	193.3, 200.1, 97.4
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	50-3.3 (3.42-3.3) *
R_{merge}	6.1 (65.3)
$I/\sigma I$	15.5 (1.4)
Completeness (%)	97.7 (96.9)
Redundancy	5.2 (3.1)
Refinement	
Resolution (Å)	50-3.3
No. reflections	56332
R_{work}/ R_{free}	22.7/26.5
No. atoms	21595
Protein	21450
Ligand/ion	145
Water	0
B-factors	
Protein	125
Ligand/ion	156
Water	n/a
R.m.s deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.47

One crystal was used.
*Highest resolution shell is shown in parenthesis.

Extended Data Table 2 | Amino acid and RNA residues

Subunit	Residues	Non-native N-terminal residues	Residues observed in structure
Rrp45	305	None	3-206, 213-301
Rrp41	246	GDPH	4-242
Rrp43	394	None	6-101,121-180, 184-191, 209-249, 271-309, 327-394
Rrp46	223	GS	1-221
Rrp42	265	GDPH	2-161, 170-264
Mtr3	250	None	5-20, 43-147, 163-248
Rrp40	240	GDPH	2-45, 54-235
Rrp4	359	GDPH	5-14, 50-145, 158-249, 277-357
Csl4	292	GDPH	5-70, 105-113, 128-162, 186-291
Rrp6 (128-685)	558	SL	128-516, 525-628
RNA	24	n.a.	19-24

LETTER

Jet acceleration of the fast molecular outflows in the Seyfert galaxy IC 5063

C. Tadhunter[1], R. Morganti[2,3], M. Rose[4], J. B. R. Oonk[2] & T. Oosterloo[2,3]

Massive outflows driven by active galactic nuclei are widely recognized to have a key role in the evolution of galaxies[1–4], by heating the ambient gas, expelling it from the nuclear regions, and thereby affecting the star-formation histories of the galaxy bulges. It has been proposed that the powerful jets of relativistic particles (such as electrons) launched by some active nuclei can both accelerate[5–7] and heat[8] the molecular gas, which often dominates the mass budgets of the outflows[5,9]. Clear evidence for this mechanism, in the form of detailed associations between the molecular gas kinematics and features in the radio-emitting jets, has however been lacking. Here we report that the warm molecular hydrogen gas in the western radio lobe of the Seyfert galaxy IC 5063 is moving at high velocities—up to about 600 kilometres per second—relative to the galaxy disk. This suggests that the molecules have been accelerated by fast shocks driven into the interstellar medium by the expanding radio jets. These results demonstrate the general feasibility of accelerating molecular outflows in fast shocks driven by active nuclei.

IC 5063 (redshift $z = 0.0113$) is a massive early-type galaxy (with stellar mass $M_* \approx 10^{11} M_\odot$, where $M_\odot$ is the mass of the Sun), which hosts both a type II Seyfert nucleus and a powerful double-lobed radio source ($P_{1.4\,\mathrm{GHz}} = 3 \times 10^{23}\,\mathrm{W\,Hz^{-1}}$). The first signs of outflows driven by the active nucleus in this object were provided by the detection of extended blue wings to the H I 21-cm absorption feature and optical [O III] emission lines at the site of the radio lobe 2.0 arcsec (0.45 kpc) to the west of its nucleus[10–12]. Subsequently, a blue wing was also detected in the CO(2–1) emission line profile of the integrated emission from the galaxy, providing evidence for molecular outflows[5]. However, the low spatial resolution of the millimetre-wavelength CO observations of this and similar objects[5–7] prevented a direct link being established between the putative molecular outflows and the relativistic jets and lobes associated with the active nucleus.

To overcome the resolution problem we have obtained deep, near-infrared long-slit spectroscopic observations of IC 5063, taken with the
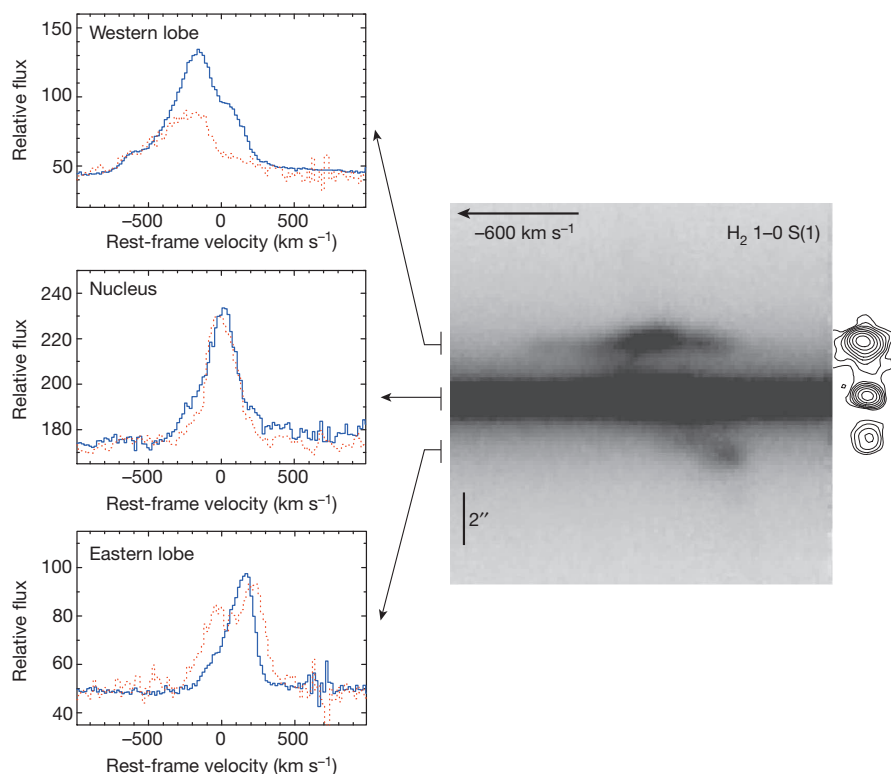
Figure 1 | Signs of extreme kinematic disturbance in the western radio lobe of IC 5063. The central panel shows a greyscale representation of our long-slit, near-infrared (K-band) spectrum of IC 5063, covering a wavelength range centred on the H_2 1–0 S(1) line. For comparison, a scaled version of the 1.4-GHz radio map of the source is presented on the right. The velocity profiles derived from spectra extracted from three spatial locations across the galaxy are presented on the left, where the solid blue lines represent the H_2 1–0 S(1) $\lambda = 2.128\,\mu m$ feature, and the dotted red lines represent the Brackett-gamma feature.

[1]Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK. [2]ASTRON, The Netherlands Institute of Radio Astronomy, PO Box 2, 7990 AA, Dwingeloo, The Netherlands. [3]Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands. [4]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

slit aligned along the axis of the extended radio lobes and jets. The observations were made in good seeing conditions (full width at half-maximum, FWHM = 0.6 arcsec) and cover the H_2 1–0 S(1) $\lambda = 2.128$ μm (change in vibrational quantum number $v = 1 - 0$; change in rotational quantum number $J = 3 - 1$) and H_2 2–1 S(2) $\lambda = 2.154$ μm ($v = 2 - 1$, $J = 4 - 2$) rotational-vibrational lines of molecular hydrogen, as well as the Brackett-gamma $\lambda = 2.166$ μm line emitted by the warm ionized hydrogen gas at the same spatial locations in the galaxy. In Fig. 1 we show a greyscale representation of the long-slit spectrum, as well as line profiles extracted for three key regions in the galaxy. Although extended molecular hydrogen emission is detected along the full 14-arcsec length of the spectroscopic slit, the surface brightness of the emission is particularly high in the regions encompassed by the lobes of the radio source (± 2 arcsec on either side of the nucleus), consistent with previous Hubble Space Telescope imaging observations[13]. Most strikingly, the kinematics of the molecular gas are highly disturbed at the position of the western radio lobe, where the H_2 1–0 S(1) $\lambda = 2.128$ μm line shows a broad, complex profile with a full width at zero intensity of FWZI $\approx$ 1,200 km s^{-1}; the H_2 emission line profile at this location is clearly broader than that of the nucleus or of the eastern radio lobe.

In Fig. 2 we show the results obtained by fitting single Gaussian profiles to the H_2 1–0 S(1) $\lambda = 2.128$ μm emission line profile at several spatial locations along the slit. From this it is clear that both the H_2 surface brightness and the linewidth peak at the position of the western radio lobe. Moreover, while the molecular gas at large radius follows the rotation curve of the extended disk of the galaxy[10,12], distortions in the radial velocity curve are apparent at the positions of the eastern and western radio lobes. Clearly, the highly disturbed emission line kinematics measured in the radio lobes cannot be explained by the normal gravitational motions of the gas in the galaxy. Therefore, these results provide clear and unambiguous evidence that the molecular gas, like the neutral H I gas[10,11], has been accelerated as a result of the interactions between the expanding radio lobes and the interstellar medium in the galaxy disk.

In terms of the comparison with the outflows detected in other phases of the interstellar medium (see Fig. 3), the H_2 line profile for the western lobe encompasses the full range of blueshifted velocities measured in the broad, H I 21-cm absorption line[10,11], but has a strong, redshifted wing that is not present in the H I feature. The latter difference can be explained by the fact that, whereas the H I absorption line samples only the gas in the foreground of the radio lobe, the H_2 emission line samples the outflowing gas moving towards and away from the observer, on the near- and far-side of the lobe, respectively. In this sense the H_2 velocity profile is similar to that of warm ionized gas, as represented by the near-infrared Brackett-gamma line, whose kinematics closely follow those of the high ionization optical emission lines (for example, [O III] $\lambda = 5,007$ Å (ref. 12). However, although the Brackett-gamma velocity profiles cover a similar velocity range to those of the H_2 line, they are different in detail (see Figs 1 and 3).

The detection of a weak H_2 2–1 S(2) $\lambda = 2.154$ μm emission line in the western lobe allows us to estimate the temperature of the molecular gas in the outflow region, since this feature has a higher excitation energy than the H_2 1–0S(1) line. The ratio between the two H_2 lines (H_2 2–1 S(2)/H_2 1–0 S(1) = 0.027 $\pm$ 0.03) is consistent with a gas temperature of $1{,}913^{+32}_{-68}$ K, assuming that the molecular gas is thermalized. Using this temperature and the spatially integrated H_2 1–0 S(1) luminosity ($L_{H_2} = (1.7 \pm 0.1) \times 10^{32}$ W), we estimate[14] a molecular hydrogen mass of $M_{H_2} = (8.2 \pm 1.2) \times 10^2$ solar masses for the western outflow region, which is several orders of magnitude lower than the H_2 mass estimated from the blueshifted CO(2–1) emission feature ($2.25 \times 10^7 < M_{H_2} < 1.29 \times 10^8$ solar masses)[5].

Our observations are consistent with a model in which the relativistic jets are expanding through the clumpy interstellar medium in the disk of the galaxy, driving fast shocks into dense molecular clouds embedded in a lower-density medium[15,16]. As the molecular gas enters the shocks it is accelerated and simultaneously heated to high temperatures ($T > 10^6$ K), ionizing the gas, and dissociating the molecules. The post-shock gas then
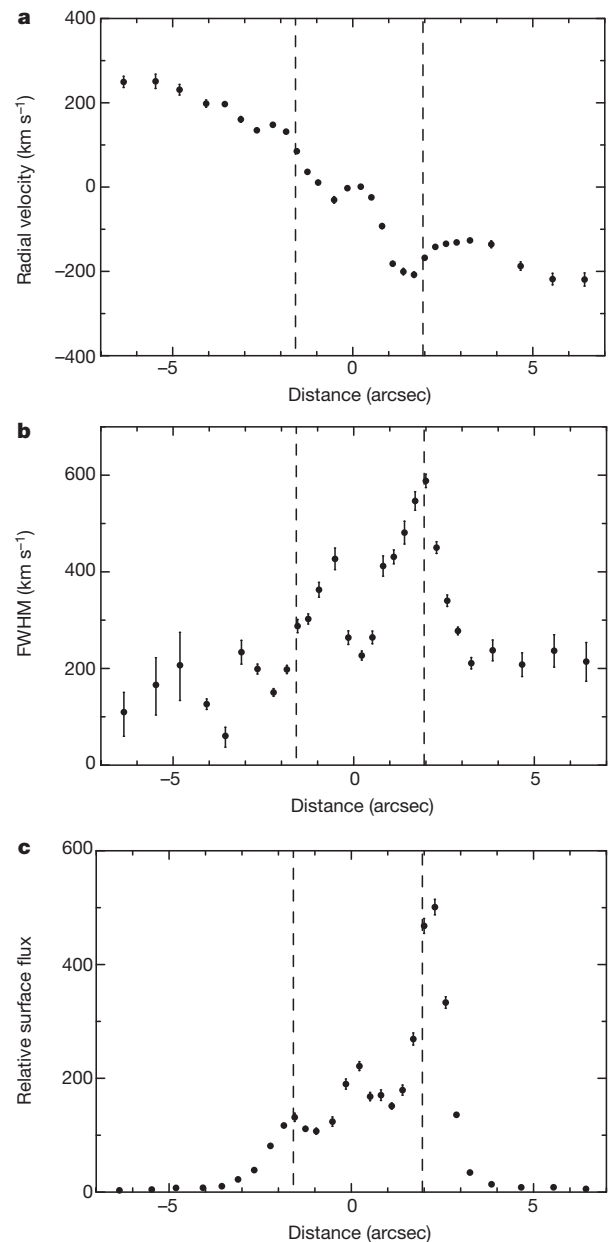
Figure 2 | Spatial variations in H_2 1–0 S(1) emission line properties along the spectroscopic slit. a, The variation in radial velocity. **b**, The variation in linewidth (FWHM). **c**, The variation in relative surface flux. Distances are measured relative to the centroid of the galaxy continuum emission, and the dashed vertical lines indicate the positions of the centroids of the radio lobes (east to the left, and west to the right). The error bars reflect the 1σ uncertainties in emission-line properties derived from single Gaussian fits to the line profiles.

cools to around 10^4 K, emitting emission lines associated with warm ionized gas (for example, Brackett-gamma) as it does so. Further cooling of the gas below 10^4 K leads to the formation of molecular hydrogen and other molecules, and the near-infrared rotational-vibrational lines of H_2 are emitted efficiently as the warm gas cools through the temperature range 5,000–1,000 K; at this stage there is also sufficient neutral hydrogen gas to allow strong absorption in the H I 21-cm line. Eventually, the molecular line emission cools the gas to low temperatures (<100 K), where it is detected through the millimetre-wavelength CO molecular lines. In this scenario, the substantial difference between the H_2 masses estimated from the near-infrared rotational-vibrational H_2 lines and the millimetre-wavelength CO lines is explained by the fact that the near-infrared H_2 lines represent a transitory phase in the warm, post-shock
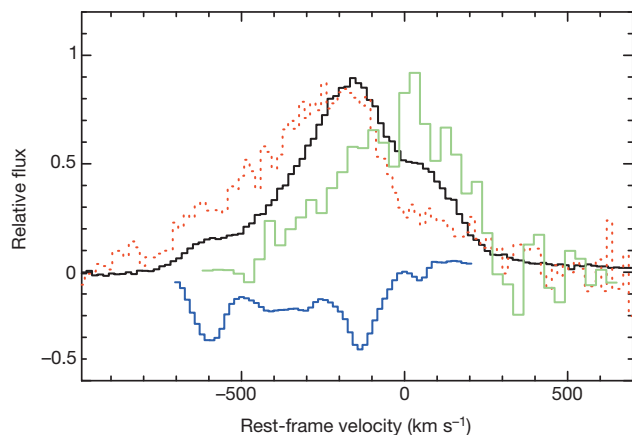
Figure 3 | The multiphase outflow in IC 5063. H_2 1–0 S(1) (black solid line) and Brackett-gamma (red dotted line) velocity profiles for the western lobe of IC 5063 are compared with the spatially integrated H I 21-cm absorption[10] (blue solid line) and CO(2–1)[5] (green solid line) velocity profiles. The flux scaling between the different profiles is arbitrary.

gas as it cools, whereas the CO lines represent the total reservoir of gas that has already cooled below 100 K.

Perhaps the greatest uncertainty with this picture concerns how the molecular hydrogen forms in the warm, post-shock gas. Given the high post-shock gas temperatures, it is likely that most, if not all, of the dust present in the precursor gas will be destroyed in the shocks. Therefore, the usual mechanism by which H_2 forms in the cool interstellar medium of the Milky Way, via catalysis on the surface of dust grains, may not be effective. In this case, it is probable that the H_2 formation is catalysed by electrons in the partially ionized cooling gas, via the intermediate formation of H^- ions[17]—the mechanism by which the molecular hydrogen formed in the first proto-stellar cores in the early Universe.

An alternative possibility is that the molecular gas has been accelerated by the slow entrainment and ablation of dense clumps of molecular gas in a hot, post-shock wind. However, in this case it is more difficult to explain the difference between the CO and near-infrared H_2 mass estimates, because we would expect the high-velocity CO- and H_2-emitting gas to have been heated to the same degree by the entrainment process. Also, we consider it unlikely that the slow entrainment process would occur in the extreme conditions of the western radio lobe, which represents the working surface of the jet.

It is also important to consider why jet-driven molecular outflows like that detected in IC 5063 appear to be rare in the general population of nearby Seyfert galaxies[18]. One possibility is that it may be necessary for the jets to collide with high-density molecular clouds in the galaxy disks for the phenomenon to be observable: perhaps only the highest-density clouds cool sufficiently quickly to avoid being destroyed by their interaction with the lower-density, post-shock wind[15]. Although this condition may be met in IC 5063, because its jets are propagating in the plane of the disk of the galaxy, in many other Seyfert galaxies the jets and the disks are not co-planar[19].

Overall, these results demonstrate the general feasibility of accelerating molecular gas in fast shocks, regardless of whether the shocks are driven by relativistic jets (as in IC 5063) or by hot, fast winds originating close to the accretion disks of active galactic nuclei[20]. Therefore they are relevant to understanding the acceleration of the massive molecular outflows that have been detected in ultraluminous infrared galaxies that contain active galactic nuclei, but lack powerful radio jets[21,22].

Although we cannot entirely rule out the alternative slow entrainment mechanism, we note that if the high-velocity molecular gas were indeed formed via cooling in the compressed, post-shock gas, the natural end point of this process would be the formation of stars. Indeed, jet-induced star formation has been invoked to explain the close alignments between the radio and optical/ultraviolet structures in high-redshift radio galaxies[23–25].

However, it has proved challenging to find definitive evidence for this mechanism, given the presence of continuum components related to the active galactic nuclei, such as scattered quasar light[26] and nebular continuum[27], which are likely to be particularly strong in powerful, high-redshift objects. At present, the best observational evidence for jet-induced star formation is provided by detailed observations of a few well-resolved radio galaxies of relatively low power in the local Universe[28,29]. Clearly, the detection of molecular hydrogen outflows in the western radio lobe of IC 5063 lends further credibility to this mechanism.

METHODS SUMMARY

The near-infrared observations of IC 5063 were taken using the medium-resolution mode of the Infrared Spectrometer and Array Camera (ISAAC) on the European Southern Observatory's Very Large Telescope, with the spectroscopic slit aligned along the radio axis (PA295). A standard ABBA nod pattern was employed, with a 20-arcsec nod throw, 3-arcsec dither box, and 300-s exposures at each position. Four repeats of the basic nod pattern resulted in a total exposure time of 4,800 s, and sky subtraction was affected by subtracting the co-aligned/co-added A and B spectra. The data were then wavelength-calibrated using the bright night-sky lines detected in the spectra, and flux-calibrated using observations of the B3V star HIP117315 taken at a similar air mass. Use of a 1.0-arcsec slit resulted in a spectral resolution of $R = 3,000$ ($100 \, km \, s^{-1}$), and the data cover a useful wavelength range of 2.104–2.230 µm, with a spatial scale of 0.146 arcsec per pixel.

Radial velocities, linewidths, and line fluxes were determined by using the STARLINK DIPSO package (http://www.starlink.rl.ac.uk/docs/sun50.htx/sun50.html) to fit single Gaussian profiles to the emission lines. All the velocities are measured relative to the rest frame of the host galaxy, as determined using the wavelength centroids of the H_2 and Brackett-gamma emission lines measured in the nucleus ($z = 0.01131 \pm 0.00004$), and the linewidths have been corrected for the instrumental profile.

At the redshift of IC 5063, 1.0 arcsec corresponds to 0.224 kpc for our assumed cosmology ($H_0 = 70 \, km \, s^{-1} \, Mpc^{-1}$; $\Omega_m = 0.3$; $\Omega_\lambda = 0.7$). Using this cosmology, the stellar mass for IC 5063 quoted in the main text was estimated from the 56-arcsec aperture K-band magnitude[30], assuming a K-band mass-to-light ratio of $(M/L_K)_\odot = 0.9$.

1. Fabian, A. Observational evidence of active galactic nuclei feedback. *Annu. Rev. Astron. Astrophys.* **50,** 455–489 (2012).
2. Silk, J. & Rees, M. Quasars and galaxy formation. *Astron. Astrophys.* **331,** L1–L4 (1998).
3. Fabian, A. The obscured growth of massive black holes. *Mon. Not. R. Astron. Soc.* **308,** L39–L43 (1999).
4. di Matteo, T. *et al.* Energy input from quasars regulates the growth and activity of black holes and their host galaxies. *Nature* **433,** 604–607 (2005).
5. Morganti, R., Frieswijk, W., Oonk, R. J. B., Osterloo, T. & Tadhunter, C. Tracing the extreme interplay between radio jets and the ISM in IC 5063. *Astron. Astrophys.* **552,** L4–L7 (2013).
6. Dasyra, K. M. & Combes, F. Cold and warm molecular gas in the outflow of 4C 12.50. *Astron. Astrophys.* **541,** L7–L11 (2012).
7. Morganti, R., Fogasy, J., Paragi, Z., Oosterloo, T. & Orienti, M. Radio jets clearing the way through a galaxy: watching feedback in action. *Science* **341,** 1082–1085 (2013).
8. Guillard, P. *et al.* Strong molecular hydrogen emission and kinematics of the multiphase gas in radio galaxies with fast jet-driven outflows. *Astrophys. J.* **747,** 95–120 (2012).
9. Alatalo, K. *et al.* Discovery of an active galactic nucleus driven molecular outflow in the early-type galaxy NGC1266. *Astrophys. J.* **735,** 88–100 (2011).
10. Morganti, R. *et al.* A radio study of the Seyfert galaxy IC 5063: evidence for fast gas outflow. *Astron. J.* **115,** 915–927 (1998).
11. Oosterloo, T. A. *et al.* A strong jet-cloud interaction in the Seyfert galaxy IC 5063: VLBI observations. *Astron. J.* **119,** 2085–2091 (2000).
12. Morganti, R., Holt, J., Saripalli, L., Oosterloo, T. A. & Tadhunter, C. N. IC 5063: AGN driven outflow of warm and cold gas. *Astron. Astrophys.* **476,** 735–743 (2007).
13. Kulkarni, V. *et al.* Unveiling the hidden nucleus of IC 5063 with NICMOS. *Astrophys. J.* **492,** L121–L124 (1998).
14. Oonk, J. B. R., Jaffe, W., Bremer, M. N. & van Weeren, R. J. The distribution and condition of the warm molecular gas in Abell 2597 and Sersic 159–03. *Mon. Not. R. Astron. Soc.* **405,** 898–932 (2010).
15. Mellema, G., Kurk, J. D. & Rottgering, H. J. A. Evolution of clouds in radio galaxy cocoons. *Astron. Astrophys.* **395,** L13–L16 (2002).
16. Wagner, A. Y., Bicknell, G. V. & Umemura, M. Driving outflows with relativistic jets and the dependence of active galactic nucleus feedback efficiency on interstellar medium inhomogeneity. *Astrophys. J.* **757,** 136–160 (2012).
17. Hollenbach, D. & McKee, C. F. Molecule formation and infrared emission in fast interstellar shocks. I. Physical processes. *Astrophys. J. Suppl. Ser.* **41,** 555–592 (1979).

18. Riffel, R. A., Storchi-Bergman, T. & Winge, C. Feeding versus feedback in AGNs from near-infrared IFU observations: the case of Mrk 79. *Mon. Not. R. Astron. Soc.* **430,** 2249–2261 (2013).
19. Schmitt, H. R., Donley, J. L., Antonucci, R. R. J., Hutchings, J. B. & Kinney, A. L. A Hubble Space Telescope survey of extended [OIII] emission in a far-infrared selected sample of Seyfert galaxies: observations. *Astrophys. J. Suppl. Ser.* **148,** 327 (2003).
20. Zubovas, K. & King, A. Galaxy-wide outflows: cold gas and star formation at high speeds. *Mon. Not. R. Astron. Soc.* **439,** 400 (2014).
21. Veilleux, S. *et al.* Fast molecular outflows in luminous galaxy mergers: evidence for quasar feedback from Herschel. *Astrophys. J.* **776,** 27–48 (2013).
22. Cicone, C. *et al.* Massive molecular outflows and evidence for feedback from CO observations. *Astron. Astrophys.* **562,** 21–46 (2014).
23. McCarthy, P. J., van Breugel, W., Spinrad, H. & Djorgovski, S. A correlation between the radio and optical morphologies of distant 3CR radio galaxies. *Astrophys. J.* **321,** L29–L33 (1987).
24. Rees, M. J. The radio/optical alignment of high-z radio galaxies—Triggering of star formation in radio lobes. *Mon. Not. R. Astron. Soc.* **239,** P1–P4 (1989).
25. Gaibler, V., Khochfar, S., Krause, M. & Silk, J. Jet-induced star formation in gas-rich galaxies. *Mon. Not. R. Astron. Soc.* **425,** 438–449 (2012).
26. di Serego Alighieri, S., Fosbury, R. A. E., Tadhunter, C. N. & Quinn, P. J. Polarized light in high-redshift radio galaxies. *Nature* **341,** 307–309 (1989).
27. Dickson, R., Tadhunter, C., Shaw, M., Clark, N. & Morganti, R. The nebular contribution to the extended UV continua of powerful radio galaxies. *Mon. Not. R. Astron. Soc.* **273,** L29–L33 (1995).
28. Croft, S. *et al.* Minkowski's object: a starburst triggered by a radio jet, revisited. *Astron. J.* **647,** 1040–1055 (2006).
29. Crockett, R. *et al.* Triggered star formation in the inner filament of Centaurus A. *Mon. Not. R. Astron. Soc.* **421,** 1603–1623 (2012).
30. Griersmith, D., Hyland, A. R. & Jones, T. J. Photometric properties of bright early-type spiral galaxies. IV—Multiaperture UBVJHK photometry for the inner/bulge regions of 65 galaxies. *Astron. J.* **1982,** 1106–1126 (1982).

Author Contributions C.T. and R.M. led the project and the scientific interpretation of the data, and C.T. wrote the text of the paper. M.R. reduced the near-infrared spectroscopic data. R.O. and T.O. contributed equally to the analysis and interpretation of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.T. (c.tadhunter@sheffield.ac.uk).

LETTER

Tracking photon jumps with repeated quantum non-demolition parity measurements

L. Sun[1]†, A. Petrenko[1], Z. Leghtas[1], B. Vlastakis[1], G. Kirchmair[1]†, K. M. Sliwa[1], A. Narla[1], M. Hatridge[1], S. Shankar[1], J. Blumoff[1], L. Frunzio[1], M. Mirrahimi[1,2], M. H. Devoret[1] & R. J. Schoelkopf[1]

Quantum error correction is required for a practical quantum computer because of the fragile nature of quantum information. In quantum error correction, information is redundantly stored in a large quantum state space and one or more observables must be monitored to reveal the occurrence of an error, without disturbing the information encoded in an unknown quantum state. Such observables, typically multi-quantum-bit parities, must correspond to a special symmetry property inherent in the encoding scheme. Measurements of these observables, or error syndromes, must also be performed in a quantum non-demolition way (projecting without further perturbing the state) and more quickly than errors occur. Previously, quantum non-demolition measurements of quantum jumps between states of well-defined energy have been performed in systems such as trapped ions[1–3], electrons[4], cavity quantum electrodynamics[5,6], nitrogen–vacancy centres[7–9] and superconducting quantum bits[10,11]. So far, however, no fast and repeated monitoring of an error syndrome has been achieved. Here we track the quantum jumps of a possible error syndrome, namely the photon number parity of a microwave cavity, by mapping this property onto an ancilla quantum bit, whose only role is to facilitate quantum state manipulation and measurement. This quantity is just the error syndrome required in a recently proposed scheme for a hardware-efficient protected quantum memory using Schrödinger cat states (quantum superpositions of different coherent states of light) in a harmonic oscillator[12]. We demonstrate the projective nature of this measurement onto a region of state space with well-defined parity by observing the collapse of a coherent state onto even or odd cat states. The measurement is fast compared with the cavity lifetime, has a high single-shot fidelity and has a 99.8 per cent probability per single measurement of leaving the parity unchanged. In combination with the deterministic encoding of quantum information in cat states realized earlier[13,14], the quantum non-demolition parity tracking that we demonstrate represents an important step towards implementing an active system that extends the lifetime of a quantum bit.

As well as being necessary in quantum error correction (QEC) and quantum information, quantum non-demolition (QND) measurements have a central role in quantum mechanics. The application of an ideal projective QND measurement yields a result corresponding to an eigenvalue of the measured operator, and projects the system onto the eigenstate associated with that eigenvalue. Moreover, the measurement must leave the system in that state, so that subsequent measurements always return the same result. The hallmark of a continuously repeated high-fidelity QND measurement is that it demonstrates a canonical thought experiment: individual quantum jumps between eigenstates are resolved in time on a single quantum system. This ideal measurement capability has been experimentally realized only in the past few decades. The jumps of a two-level system (quantum bit, or qubit) between its energy eigenstates were first observed for single trapped ions[1–3], and later in single nitrogen–vacancy centres in diamond[7–9]. The jumps of an oscillator between

eigenstates with different numbers of excitations (Fock states), were first observed for the motion of an electron in a Penning trap[4]. More recently, the observation of quantum jumps of light in cavity quantum electrodynamics[5,6] (QED), where the number of microwave photons in a cavity is probed with Rydberg atoms, has enabled a range of new experiments in quantum feedback and control[15,16].

An analogous system to cavity QED is the combination of microwave photons in a superconducting resonator with superconducting qubits, known as circuit QED[17]. The strong dispersive interaction of a qubit and a photon, as in Rydberg-atom cavity QED, allows either the qubit or the cavity to act as a QND probe of the other component. With the advent of quantum-limited parametric amplifiers[18,19], measurement techniques for superconducting devices have rapidly advanced. For instance, the frequency shift of a cavity has recently been used to observe the quantum jumps of a qubit between energy eigenstates[10,11]. So far, however, there have been no observations of jumps for the cavity field in circuit QED. A recent paper measured a different quantity, the parity of two qubits, in a step towards the conventional approach of QEC[20]. However, that work did not present real-time tracking of the jumps due to the natural error rate of that quantity.

In this work, we use the dispersive qubit–cavity interaction of circuit QED to observe the jumps of photon number parity. Importantly, these jumps reveal the loss of individual photons without projecting the system onto a state of definite number or energy, but rather into an eigenspace of even or odd photon number. This characteristic is a crucial requirement for future applications in quantum information, where the parity measurement serves as the error syndrome for correcting a quantum memory. Even in the presence of rapidly repeated measurements, the smooth decay of the ensemble-averaged parity is largely unperturbed. However, when individual time records of the measurement are examined, the parity is observed to take on only the extremal values, ± 1, indicating the projective nature of each individual measurement. On examining the statistics of the jumps recorded over many trajectories, we find excellent agreement with a numerical simulation, suggesting that 85% of the jumps for states with an average photon number $\bar{n} = 4$ are faithfully detected (see Methods section on photon jump statistics). When selecting on the outcome of a single parity measurement, we observe, by Wigner tomography[21], the creation of cat states with $\bar{n}$ up to 4.

In our experiment, we use a three-dimensional circuit QED architecture[22] with a single 'vertical' superconducting transmon qubit (the qualifier 'vertical' indicates that the dominant electric field is perpendicular to the film plane) coupled to two waveguide cavities[14,23], as shown in Fig. 1a. Our qubit has a transition frequency of $\omega_q/2\pi = 5.938$ GHz, an energy relaxation time of $T_1 = 8$ μs and a Ramsey time of $T_2^* = 5$ μs. The high-frequency cavity, with a resonant frequency of $\omega_m/2\pi = 8.174$ GHz and a lifetime of 30 ns, serves only as a fast readout of the qubit state. To perform a high-fidelity single-shot dispersive readout of the qubit, we use a Josephson bifurcation amplifier (JBA) operating in a double-pumped

[1]Departments of Applied Physics and Physics, Yale University, New Haven, Connecticut 06511, USA. [2]INRIA Paris-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France. †Present addresses: Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China (L.S.); Institut für Experimentalphysik, Universität Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria and Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Otto-Hittmair-Platz 1, A-6020 Innsbruck, Austria (G.K.).
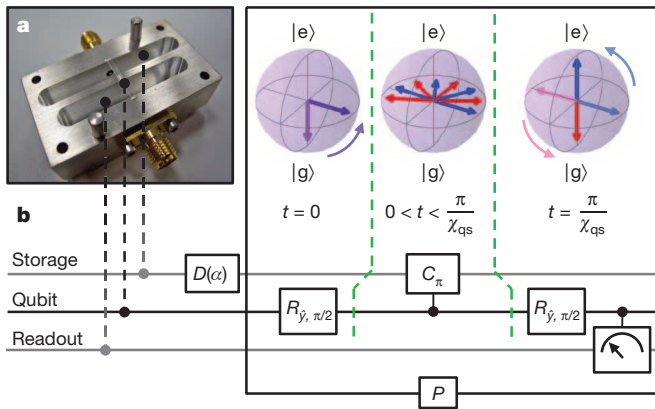
Figure 1 | Experimental device and parity measurement protocol of a photon state. a, Bottom half of the device containing a 'vertical' transmon qubit located in a trench and coupled to two waveguide cavities. The low-frequency cavity, with $\omega_s/2\pi = 7.216$ GHz and a lifetime of $\tau_0 = 55$ μs, is used to store and manipulate quantum states. The high-frequency cavity, with $\omega_m/2\pi = 8.174$ GHz and a lifetime of 30 ns, allows for fast readout of the qubit. **b**, Protocol (P) for measuring the parity of the storage cavity field. After an initial displacement of cavity vacuum $D(\alpha)|0\rangle = |\alpha\rangle$ to create a coherent state with a complex amplitude α, a Ramsey-type measurement is performed. It consists of two π/2-pulses separated by $t = \pi/\chi_{qs}$ (during which a controlled-phase gate $C_\pi = I\otimes|g\rangle\langle g| + e^{i\pi a^\dagger a}\otimes|e\rangle\langle e|$ is realized), followed by a projective measurement of the qubit, where χ_{qs} is the dispersive interaction between the qubit and the storage cavity. In this schematic, with the qubit initially in the ground state, $|g\rangle$, the Ramsey-type measurement maps the even and odd photon states onto the $|e\rangle$ and $|g\rangle$ states of the qubit, respectively. A subsequent projective measurement indicates the cavity state parity. The second π/2-pulse can be either $R_{\hat{y}, -\pi/2}$ or $R_{\hat{y}, \pi/2}$, simply switching the interpretation of the result of the qubit measurement.

mode[24,25] as the first stage of amplification. The low-frequency cavity, with a resonant frequency of $\omega_s/2\pi = 7.216$ GHz and a lifetime of $\tau_0 = 55$ μs, stores the photon states which are measured and manipulated. Exploiting the nonlinearities induced in both resonators, we use the transmon qubit to track the parity of the storage cavity state. For simplicity, we will refer to the storage cavity as the 'cavity' henceforth.

The qubit and cavity are in the regime of strong dispersive coupling, which can be described by the Hamiltonian

$$H/\hbar = \omega_q|e\rangle\langle e| + \left(\omega_s - \chi_{qs}|e\rangle\langle e|\right)a^\dagger a$$

where a and $a^\dagger$ are the annihilation and creation operators, respectively, $|e\rangle$ is the excited state of the qubit and $\chi_{qs}/2\pi = 1.789$ MHz is the qubit-state-dependent frequency shift of the cavity. The readout cavity has been neglected because it remains in the ground state while the system evolves. The interaction between the qubit and the cavity entangles qubit and photon. In the rotating frame of the cavity, Fock states associated with the qubit in the excited state acquire a phase $\Phi = a^\dagger a \chi_{qs} t$ proportional to their photon number[26]. By waiting for $t = \pi/\chi_{qs}$, one can realize a controlled-phase gate $C_\pi = I\otimes|g\rangle\langle g| + e^{i\pi a^\dagger a}\otimes|e\rangle\langle e|$, where $|g\rangle$ is the ground state of the qubit, adding a phase shift of π per photon to the cavity state conditioned on the qubit state[14,27]. Therefore, C_π can be inserted between two π/2-pulses on the qubit in a Ramsey-type measurement to map the photon parity of any cavity state onto the qubit (black enclosure labelled 'P' in Fig. 1b). The result of a qubit measurement after the second π/2-pulse, together with prior knowledge of the initial qubit state, indicates whether the number of photons in the cavity is even or odd, but reveals nothing about the actual value of the photon number.

The creation of cat states is a natural consequence of a parity measurement on a coherent state $|\alpha\rangle$ (α is a complex amplitude) because the phase cat states defined by $\mathcal{N}_\pm(|\alpha\rangle \pm |-\alpha\rangle)$, with $\mathcal{N}_\pm = 1/\sqrt{2(1 \pm e^{-2|\alpha|^2})}$, are eigenstates of the parity operator $e^{i\pi a^\dagger a}$ (refs 28, 29). After applying a microwave pulse at frequency ω_s to the cavity, initially in vacuum,

to create a coherent state $|\alpha\rangle$ with the qubit initially in $|g\rangle$, we use the parity protocol to take $|\alpha\rangle(|g\rangle + |e\rangle)/\sqrt{2}$ after the first π/2-pulse to $[\mathcal{N}_-(|\alpha\rangle - |-\alpha\rangle)|g\rangle + \mathcal{N}_+(|\alpha\rangle + |-\alpha\rangle)|e\rangle]/\sqrt{2}$ after the second pulse, at which point the parity of the cavity state is entangled with the state of the qubit. Detection of the qubit state using the readout cavity then projects the storage cavity onto one of the two cat states. To confirm the non-classical properties of these states, we perform Wigner tomography of the cavity after a single parity measurement for an initial coherent state of displacement $|\alpha| = 2$ ($\bar{n} = 4$). Post-selecting on the ground or excited qubit states to obtain the odd or even cats (Fig. 2a, b), respectively, we see the interference patterns that are the signature of quantum behaviour. The overlap between the measured Wigner function and that of an ideal cat state gives a fidelity of $F = 83\%$ for the odd cat state. Figure 2c shows the Wigner function without post-selection (tracing over qubit states). Fringes in the Wigner function almost completely disappear, as expected, and we obtain the statistical mixture of even and odd states. The weak visibility of the fringes comes from the slightly lower fidelity of the even cat state, wherein the qubit ends up in the $|e\rangle$ state, which is more susceptible to qubit relaxation. Figure 2d shows the normalized difference between the two cat states to emphasize the interference fringes. The high contrast between even and odd cat states is a central requirement in implementing a recently proposed QEC scheme[12], where these form the code and error spaces, respectively.

Because the loss of a single photon changes the parity of a cat state, monitoring parity repeatedly in real time allows us to track photon jumps of our cavity. Here we note that to interpret the result of a single parity measurement we must know the state of the qubit before the first π/2-pulse. In other words, it is the correlation of the qubit states before and after the parity measurement (a pattern of oscillation between $|g\rangle$ and $|e\rangle$ versus a constant pattern remaining in either $|g\rangle$ or $|e\rangle$) that reveals the photon state parity. For the following data we have chosen $R_{\hat{y}, -\pi/2}$ as the second qubit pulse, instead of $R_{\hat{y}, \pi/2}$, to maintain a constant pattern when the cavity is in the even parity state. Apart from reversing which pattern we assign to be even and which we assign to be odd, this change makes no difference. Figure 3a shows the measurement protocol and Fig. 3b–e shows typical 400 μs single-shot traces. The initial displacement is $|\alpha| = 1.0$ and the repetition interval of the parity measurements is 1 μs, which is much smaller than the average photon lifetime, $\tau_0 = 55$ μs,
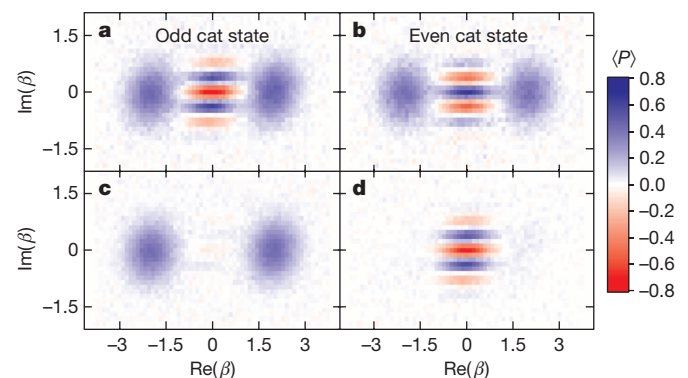
Figure 2 | Ensemble-averaged Wigner functions of cat states in the cavity created by single-shot parity measurements of an initial coherent state in the cavity. The Wigner functions are mapped out with varying displacements β and a measurement of the mean photon parity $\langle P \rangle$ (ref. 21). Here we follow the protocol depicted in Fig. 1b, using $R_{\hat{y}, \pi/2}$ as the second pulse. The qubit is always initialized to the $|g\rangle$ state through post-selection on an initial measurement. **a**, Odd cat state by post-selection on the $|g\rangle$ state as the result of the parity measurements. **b**, Even cat state by post-selection on the $|e\rangle$ state. **c**, No post-selection of the parity measurement, thus tracing over the qubit state. Fringes almost disappear, indicating a mixture of two coherent states. **d**, The normalized difference (data in **a** minus data in **b**, all divided by two), or the expectation of the parity weighted by $\langle \sigma_z \rangle$ of the ancilla, emphasizing the interference fringes.
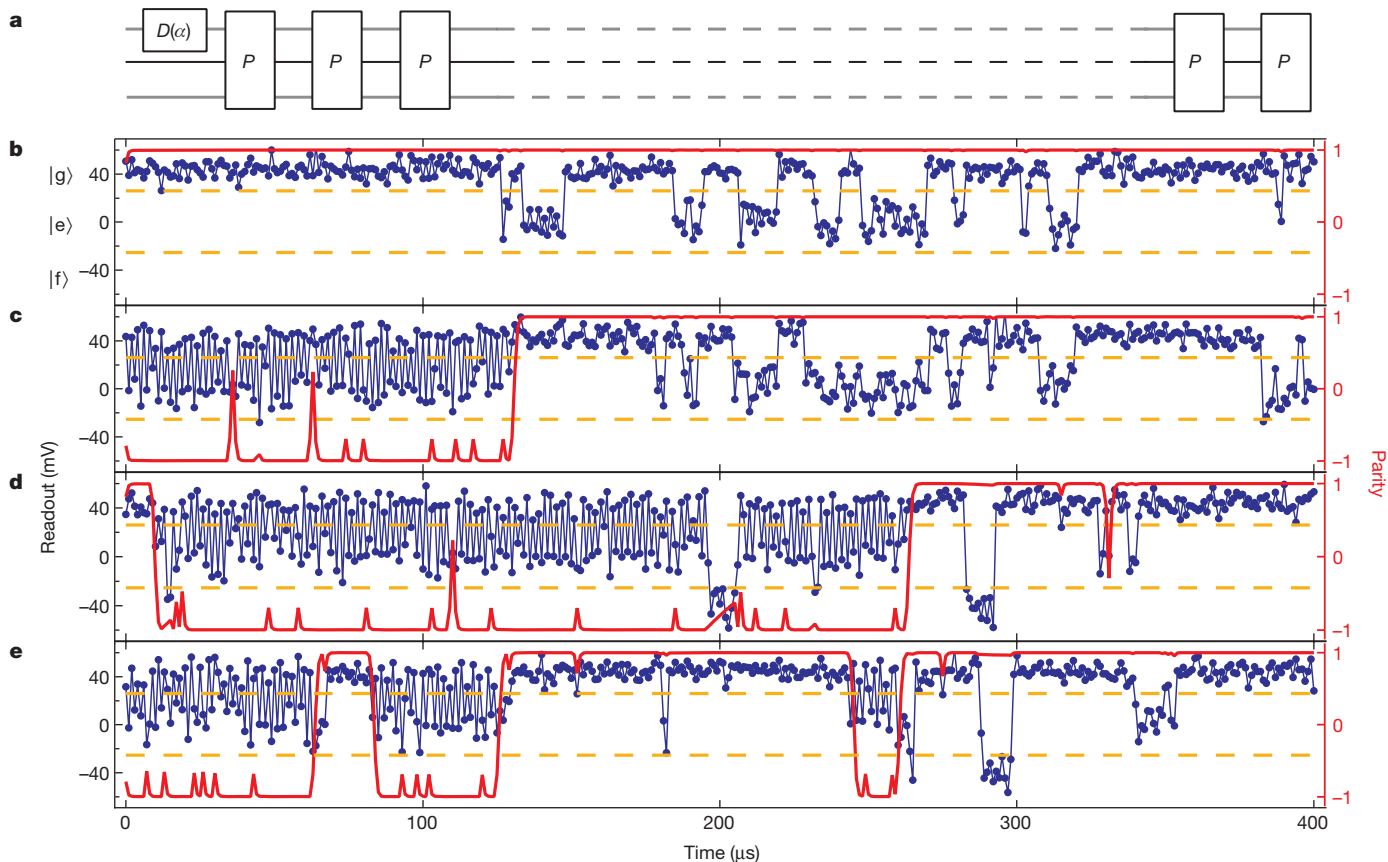
Figure 3 | Typical repeated single-shot parity measurement traces revealing photon jumps in real time. Horizontal dashed orange lines delineate the thresholds to distinguish $|g\rangle$, $|e\rangle$ and higher excited states of the qubit, denoted as $|f\rangle$. The red traces show the quantum filter that best estimates the parity at every point. The filter has a finite response time and thus does not trust that a brief change in the measurement pattern corresponds to an actual parity jump. **a**, In this protocol we switch the sign of the second pulse, using $R_{\hat{y},-\pi/2}$ instead of $R_{\hat{y},\pi/2}$. The repetition time of the parity measurement is 1 μs, and the traces in **b**–**e** all have an initial displacement of $|\alpha| = 1$. **b**, For the most part, the correlation between neighbouring measurements is positive, indicating

an even-parity state for the whole 400 μs. The changes in the qubit state between 120 μs and 320 μs are probably due to qubit decoherence during the parity measurement. **c**, One parity jump is observed by the change in the measurement pattern (oscillating versus constant) at about 130 μs. **d**, Two parity jumps are recorded at about 10 μs and then again at 260 μs. The change of pattern at about 200 μs is a result of the qubit leaving the computational space for higher excited states, a feature that disables the parity measurement until the qubit returns to either $|g\rangle$ or $|e\rangle$. **e**, A trace with all features described above included. In this particular trajectory, the filter can clearly resolve five photon jump events.

obtained from a free time evolution measurement of the parity of a coherent state (see Methods section on experimental set-up). We observe a range of photon jump statistics, from quiet traces that last for hundreds of microseconds with no apparent changes in parity, to those that have as many as five jumps. The clear dichotomy between the patterns in our traces indicates that, although the measurements are susceptible to qubit decoherence, as evidenced by intermittent, brief changes in measurement correlations and excitations to higher qubit states, they nonetheless exhibit a strong sensitivity to single-photon jump events.

When analysing these single-shot traces, to mitigate the effects due to qubit decoherence, excitation to qubit states higher than $|e\rangle$ (denoted as $|f\rangle$) and other imperfections in the qubit readout in extracting the parity, we have applied a quantum filter that best estimates the photon state parity (see Methods section on the quantum filter). We note that the output of the quantum filter depends on the entire previous parity trajectory. Figure 3b–e shows traces with the parity estimator calculated from the quantum filter, in red. The parity estimator is clearly much less sensitive to qubit decoherence and $|f\rangle$ states. Although our single parity readout fidelity is 80%, owing to the smoothing effect of the quantum filter we actually can achieve nearly unity detection sensitivity of single-photon jumps. However, given one jump, the probability of there being a second jump within the response time of the filter (∼2 μs) is 4% for $\bar{n} = 1$ (or 15% for $\bar{n} = 4$), which limits our overall detection sensitivity over an entire trajectory (see Methods section on photon jump statistics).

The repeated parity measurements shown above constitute just a single point, the origin, in the Wigner functions of the even and odd cat states (Fig. 2a, b). Thus, crucially, a parity measurement acquires no information about the phase of the cat states. Consequently, one could encode quantum information onto the computational bases $|0\rangle_L = \mathcal{N}_+(|\alpha\rangle + |-\alpha\rangle)$ and $|1\rangle_L = \mathcal{N}_+(|i\alpha\rangle + |-i\alpha\rangle)$, and any subsequent parity measurements would make no distinction between the two. The loss of a single photon will change the code space spanned by $|0\rangle_L$ and $|1\rangle_L$ into the error space spanned by $|\bar{0}\rangle_L = \mathcal{N}_-(|\alpha\rangle - |-\alpha\rangle)$ and $|\bar{1}\rangle_L = \mathcal{N}_-(|i\alpha\rangle - |-i\alpha\rangle)$ with a different parity. This error syndrome can thus be extracted by the parity measurement demonstrated here, but without gaining any knowledge of the information encoded in the cat states, as required by QEC.

The degree to which the measurements are QND can be determined by examining the decay rate for the parity of a coherent state with different measurement cadences. We extract the total decay rate of the parity (τ_{tot}), from the ensemble-averaged parity dynamics obtained with the quantum filter (Fig. 4). This total decay rate is well modelled by the parallel combination of the free decay time ($\tau_0 = 55$ μs) plus a constant demolition probability $P_D = 0.002$ per measurement interval τ_i, as shown by the fit in the inset of Fig. 4. In other words, a single parity measurement is 99.8% QND, leaving the parity of the cavity state largely unperturbed.

Several improvements and further investigations will be required to realize a truly robust error-corrected quantum memory. The probability of missing a photon jump, as a result of the finite measurement rate
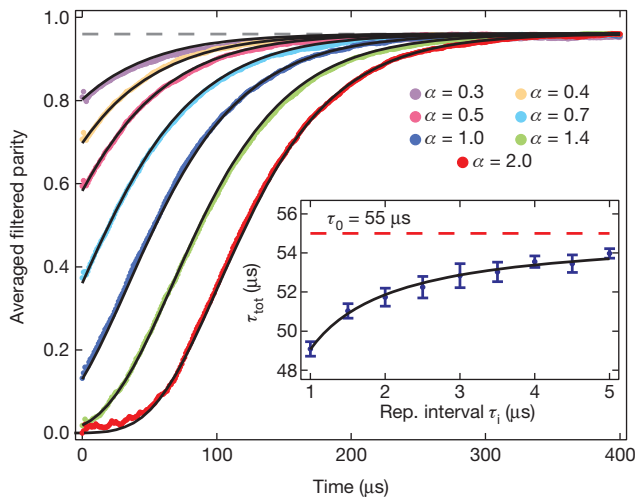
Figure 4 | Ensemble average of filtered single-shot parity traces for different cavity displacements α. The $|\alpha| = 1$ curve is the average of 20,000 traces, the examples of which are shown in red in Fig. 3b–e. Note that there is no post-selection on the initial parity of the cat state. Parity measurements are repeated every $\tau_i = 1\,\mu s$. The parity of an initial coherent state is expected to evolve as $P = e^{-2|\alpha|^2 e^{-t/\tau_{tot}}/(1+2n_{th})}/(1 + 2n_{th})$, where τ_{tot} is the photon decay time and n_{th} is the background photon number in the cavity (see Methods section on the quantum filter). Given $n_{th} = 0.02$ (from an independent measurement; see Methods section on the experimental set-up), the average parity is expected to saturate at $1/(1 + 2n_{th}) = 0.96$ (the dashed horizontal line). Solid lines are fitted theoretical curves for coherent states where n_{th} and the corresponding $|\alpha|$ are fixed, leaving τ_{tot} as the only free parameter. Here the value of $\tau_{tot} = 49\,\mu s$ gained from a global fit closely matches the lifetime $\tau_0 = 55\,\mu s$ obtained from free evolution, and each trace indeed saturates at $P = 0.96$. This indicates that repeated parity measurements do not significantly perturb the cavity state, inducing neither extra jumps nor an increase in cavity thermal population. Inset, extracted time constants as functions of different parity measurement repetition intervals. The error bars indicate one standard deviation calculated from nine measurements, as in the main panel. The decay time τ_{tot} is modelled as $1/\tau_{tot} = 1/\tau_0 + P_D/\tau_i$, where P_D is the probability of inducing an extra parity change. A fit (solid line) gives $P_D = 2 \times 10^{-3}$, indicating that a single parity measurement is 99.8% QND.

per cavity lifetime, would be greatly reduced if longer-lived cavities were used[30]. It is a necessary, but not sufficient, condition that the measurement itself, as shown here, is highly QND and unlikely to induce photon jumps. An additional requirement is that the measurement does not destroy the actual quantum information stored in the cat states. Dephasing of the cavity state will be non-negligible because the current realization is not yet robust against qubit decay or excitation (see Methods section on quantifying the parity-tracking performance). Increasing qubit and cavity lifetimes, and further characterizing these types of error processes, are important next steps. Nonetheless, we estimate that when combined with an optimized measurement strategy, the current level of performance could already allow extension of the average lifetime of an encoded cat state ($\tau_{tot}/\bar{n}$) by a factor of two (see Methods section on parity-tracking performance). There are several other theoretical architectures for QEC, some of which have already been implemented experimentally. However, these experiments have not yet reached the level where they can suppress the naturally occurring errors. Understanding the role of the measurement process and its imperfections in all of these approaches remains an important topic for current and future work.

We have demonstrated the real-time tracking of jumps in the photon number parity in circuit QED. Significantly, this quantity differs from previous observations of quantum jumps between energy levels. Rather, it projects the system into a degenerate subspace, and can therefore serve as an error syndrome for QEC. We show that the parity measurement is highly QND and that it has a high fidelity and cadence compared with the cavity lifetime. These advances in the measurement capabilities should enable further progress in quantum information.

METHODS SUMMARY

Measurements are performed in a cryogen-free dilution refrigerator with a base temperature of about 10 mK. The 'vertical' transmon qubit is fabricated on a c-plane sapphire (Al_2O_3) substrate with a double-angle evaporation of aluminium after a single electron-beam lithography step. The state-dependent frequency shift between the qubit and the readout cavity is $\chi_{qr}/2\pi = 0.930$ MHz, which is not optimized for the best signal-to-noise ratio. The background photon number $n_{th} = 0.02$ and the displacement α are calibrated on the basis of the Poisson distribution of photon numbers in the storage cavity (Extended Data Fig. 2). The qubit readout fidelity is about 90%, mainly limited by the short coherence times of the qubit. The quantum filter used as the best-parity estimator consists of two steps: a time evolution of the density matrix taking into account the cavity decoherence, and a modification of the density matrix based on the current measurement result. The effectiveness of this filter is confirmed by the good agreement between extracted numbers of jumps from the parity estimator during 500 μs repeated parity measurements and a numerical simulation (Extended Data Fig. 9). The Wigner tomography fidelity is limited by qubit T_1, T_ϕ, photon jumps and detection inaccuracy. Among these factors, only qubit T_1 and missing of fast photon jumps can lead to the decay of the cat states. The high-QND nature of the parity measurement allows the incorrect parity measurement result due to qubit T_ϕ and detection inaccuracy to be removed by performing repeated parity measurements and taking a majority voting. Slow photon jumps can be tracked as demonstrated in the manuscript, and the resulting phase errors of the cat states can also be completely corrected. Repeated parity tracking can thus enhance the lifetime of the information encoded in cat states (Methods section on parity-tracking performance).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Bergquist, J. C., Hulet, R. G., Itano, W. M. & Wineland, D. J. Observation of quantum jumps in a single atom. *Phys. Rev. Lett.* **57**, 1699–1702 (1986).
2. Sauter, T., Neuhauser, W., Blatt, R. & Toschek, P. E. Observation of quantum jumps. *Phys. Rev. Lett.* **57**, 1696–1698 (1986).
3. Nagourney, W., Sandberg, J. & Dehmelt, H. Shelved optical electron amplifier: observation of quantum jumps. *Phys. Rev. Lett.* **56**, 2797–2799 (1986).
4. Peil, S. & Gabrielse, G. Observing the quantum limit of an electron cyclotron: QND measurements of quantum jumps between Fock states. *Phys. Rev. Lett.* **83**, 1287–1290 (1999).
5. Gleyzes, S. *et al.* Quantum jumps of light recording the birth and death of a photon in a cavity. *Nature* **446**, 297–300 (2007).
6. Guerlin, C. *et al.* Progressive field-state collapse and quantum non-demolition photon counting. *Nature* **448**, 889–893 (2007).
7. Jelezko, F. *et al.* Single spin states in a defect center resolved by optical spectroscopy. *Appl. Phys. Lett.* **81**, 2160–2162 (2002).
8. Neumann, P. *et al.* Single-shot readout of a single nuclear spin. *Science* **329**, 542–544 (2010).
9. Robledo, L. *et al.* High-fidelity projective read-out of a solid-state spin quantum register. *Nature* **477**, 574–578 (2011).
10. Vijay, R., Slichter, D. H. & Siddiqi, I. Observation of quantum jumps in a superconducting artificial atom. *Phys. Rev. Lett.* **106**, 110502 (2011).
11. Hatridge, M. *et al.* Quantum back-action of an individual variable-strength measurement. *Science* **339**, 178–181 (2013).
12. Leghtas, Z. *et al.* Hardware-efficient autonomous quantum error correction. *Phys. Rev. Lett.* **111**, 120501 (2013).
13. Leghtas, Z. *et al.* Deterministic protocol for mapping a qubit to coherent state superposition in a cavity. *Phys. Rev. A* **87**, 042315 (2013).
14. Vlastakis, B. *et al.* Deterministically encoding quantum information using 100-photon Schrödinger cat states. *Science* **342**, 607–610 (2013).
15. Deléglise, S. *et al.* Reconstruction of non-classical cavity field states with snapshots of their decoherence. *Nature* **455**, 510–514 (2008).
16. Sayrin, C. *et al.* Real-time quantum feedback prepares and stabilizes photon number states. *Nature* **477**, 73–77 (2011).
17. Devoret, M. H. & Schoelkopf, R. J. Superconducting circuits for quantum information: an outlook. *Science* **339**, 1169–1174 (2013).
18. Castellanos-Beltran, M. A., Irwin, K. D., Hilton, G. C., Vale, L. R. & Lehnert, K. W. Amplification and squeezing of quantum noise with a tunable Josephson metamaterial. *Nature Phys.* **4**, 929–931 (2008).
19. Bergeal, N. *et al.* Phase-preserving amplification near the quantum limit with a Josephson ring modulator. *Nature* **465**, 64–68 (2010).
20. Ristè, D. *et al.* Deterministic entanglement of superconducting qubits by parity measurement and feedback. *Nature* **502**, 350–354 (2013).
21. Haroche, S. & Raimond, J. M. *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford Univ. Press, 2006).
22. Paik, H. *et al.* Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107**, 240501 (2011).
23. Kirchmair, G. *et al.* Observation of quantum state collapse and revival due to the single-photon Kerr effect. *Nature* **495**, 205–209 (2013).

24. Kamal, A., Marblestone, A. & Devoret, M. H. Signal-to-pump back action and self-oscillation in double-pump Josephson parametric amplifier. *Phys. Rev. B* **79,** 184301 (2009).
25. Murch, K. W., Weber, S. J., Macklin, C. & Siddiqi, I. Observing single quantum trajectories of a superconducting quantum bit. *Nature* **502,** 211–214 (2013).
26. Schuster, D. I. *et al.* Resolving photon number states in a superconducting circuit. *Nature* **445,** 515–518 (2007).
27. Bertet, P. *et al.* Direct measurement of the Wigner function of a one-photon Fock state in a cavity. *Phys. Rev. Lett.* **89,** 200402 (2002).
28. Brune, M., Haroche, S., Raimond, J. M., Davidovich, L. & Zagury, N. Manipulation of photons in a cavity by dispersive atom-field coupling: Quantum-nondemolition measurements and generation of ''Schrödinger'' cat states. *Phys. Rev. A* **45,** 5193–5214 (1992).
29. Brune, M. *et al.* Observing the progressive decoherence of the ''meter'' in a quantum measurement. *Phys. Rev. Lett.* **77,** 4887–4890 (1996).
30. Reagor, M. *et al.* Ten milliseconds for aluminum cavities in the quantum regime. *Appl. Phys. Lett.* **102,** 192604 (2013).

METHODS

Experiment set-up, device parameters and qubit readout properties. Our measurements are performed in a cryogen-free dilution refrigerator with a base temperature of about 10 mK. Extended Data Fig. 1 shows the schematic of the measurement set-up. A Josephson bifurcation amplifier[31,32] (JBA) operating in a double-pumped mode[24,25] is used as the first stage of amplification between the readout cavity output and the high-electron-mobility transistor (HEMT), allowing for a high-fidelity single-shot dispersive readout of the qubit state. We typically operate the JBA in the saturated regime with about 20 readout photons for a better signal-to-noise ratio.

The 'vertical' transmon qubit is fabricated on a c-plane sapphire (Al_2O_3) substrate with a double-angle evaporation of aluminium after a single electron-beam lithography step. We use the word 'vertical' to specify that the dominant electric field is perpendicular to the film plane. The qubit has a transition frequency $\omega_q/2\pi = 5.938$ GHz with an anharmonicity of $\alpha_q/2\pi = (\omega_{ge} - \omega_{ef})/2\pi = 240$ MHz, an energy relaxation time of $T_1 = 8$ μs and a Ramsey time of $T_2^* = 5$ μs. These coherence times are short compared with those in a regular three-dimensional cavity[22], but are comparable to those reported in refs 14, 23. Exploring the exact sources of limitation is an on-going research subject. Even at the lowest base temperature, the qubit steady state is measured to be a thermal mixture of about 86% ground state $|g\rangle$, 11% excited state $|e\rangle$ and 3% states higher than $|e\rangle$, denoted as $|f\rangle$. These excitations of the qubit could come from stray infrared photons leaking into the cavity, although the exact source remains unknown.

The qubit serves as an ancilla and provides the necessary nonlinearity for the manipulation of coherent states in the storage cavity. Both the storage and readout cavities are made of aluminium alloy 6061. The state-dependent frequency shifts between the qubit and the storage and readout cavities are $\chi_{qs}/2\pi = 1.789$ MHz and $\chi_{qr}/2\pi = 0.930$ MHz, respectively. For simplicity, we will refer to the storage cavity as the 'cavity' henceforth. The inset of Extended Data Fig. 2 shows the 'number splitting peaks' of the qubit due to different photon numbers in the cavity, which is displaced with a 10 ns square pulse right before the spectroscopy measurement. A second-order polynomial fit $\chi(n) = -\chi_{qs}n + \chi'_{qs}n^2$, where n is the peak number, gives a nonlinear correction to the dispersive shift[14] $\chi'_{qs}/2\pi = 1.9 \pm 0.1$ kHz which is small enough to be neglected in the cavity dynamics. Extended Data Fig. 2 shows that the probabilities of the first eight Fock states $n = 0, 1, 2, …, 7$ as functions of displacement amplitude $|\alpha|$ are in excellent agreement with a Poisson distribution, indicating good control of the coherent state in the cavity. We scale the x axis from the voltage amplitude of the displacement pulse applied from an arbitrary waveform generator and use this scaling as a calibration. There is a small residual amplitude for the $n = 1$ peak even with no displacement (point near origin), allowing us to infer that there is a background photon population $n_{th} = 0.02$ in the cavity. The lifetime of the cavity is characterized by measuring a free parity evolution of a coherent state, as shown in Extended Data Fig. 3, which is nearly identical to Fig. 4. A global fit gives a time constant $\tau_0 = 55$ μs. We also perform another experiment by introducing a small average photon number ($\bar{n} \approx 0.1$) in the cavity and monitoring the exponential decay of the $n = 1$ qubit peak. This experiment gives an identical τ_0. However, the free parity evolution of a coherent state can have much better signal-to-noise ratio because large $\bar{n}$ (as large as 4) can be introduced in the cavity.

To perform a good parity measurement, the $\pi/2$-pulses $R_{\hat{y}, \pm \pi/2}$ should equally cover as many number splitting peaks as possible without significantly exciting the $|f\rangle$ state. We choose a Gaussian envelope pulse truncated to $4\sigma = 8$ ns ($\sigma_f = 80$ MHz) for a good compromise. Extended Data Fig. 4 shows the effectiveness of those $R_{\hat{y}, \pm \pi/2}$ pulses as a function of $\bar{n}$ in the cavity. The curvature for $\bar{n} > 4$ is due to the finite bandwidth of those pulses in the frequency domain.

We have adjusted the phase between the JBA readout signal and the pump such that $|g\rangle$, $|e\rangle$ and $|f\rangle$ states can be distinguished with optimal contrast. Extended Data Fig. 5a shows the histogram of the qubit readout for the parity protocol used in repeated single-shot traces in Fig. 3. The histogram is clearly trimodal. Thresholds between $|g\rangle$ and $|e\rangle$, and between $|e\rangle$ and $|f\rangle$ states, have been chosen to digitize the readout signal to $+1$, -1 and 0 for $|g\rangle$, $|e\rangle$ and $|f\rangle$ states, respectively. We assign a zero to the $|f\rangle$ states to indicate a 'failed' measurement with no useful information about the parity. These $|f\rangle$ states can be fixed with a field-programmable gate array applying proper pulses to drive the qubit back to either $|g\rangle$ or $|e\rangle$ in real time. Extended Data Fig. 5e shows the basic qubit readout properties with the cavity left in vacuum. The $|g\rangle$ state is prepared through a post-selection of an initial qubit measurement, while $|e\rangle$ and $|f\rangle$ are prepared by properly pulsing the selected $|g\rangle$ state (see Extended Data Fig. 5b–d for the pulse sequences). The loss of fidelity predominantly comes from the T_1 process during both the waiting time of the initialization measurement (500 ns) and the qubit readout time (300 ns).

Wigner tomography fidelity and parity readout fidelity. We emphasize that it is the correlation C_t of the qubit states before and after the parity measurement that reveals the photon state parity. Extended Data Fig. 6a shows the parity readout properties of our system for the protocol ($R_{\hat{y}, -\pi/2}$ as the second qubit pulse) used in the single-shot traces in Fig. 3. The loss of fidelity of the parity measurement

mainly comes from qubit decoherence processes during the parity measurement (discussed later). Conditional probabilities P($+1$|even), P($+1$|odd), P(-1|even), P(-1|odd), P(0|even) and P(0|odd) are time-independent probabilities that have positive, negative and zero correlations (as indicated) between the digitized qubit readouts before and after a parity measurement for a given even or odd state. However, a pure even or odd state cannot be prepared easily in our system owing to the finite thermal population of the cavity, which is small but can still introduce systematic errors. We determine P(± 1,0|even/odd) by post-selecting the cases with five consecutive identical parity results, which give the photon state parity with good confidence, and then performing a histogram on the sixth parity measurement (Extended Data Fig. 6b).

Extended Data Fig. 6c shows the pulse sequence for producing the cat states and the Wigner tomography shown in Fig. 2. The protocol starts with a post-selection of the $|g\rangle$ state of the qubit after an initial qubit measurement M_1. A parity measurement is performed immediately after a storage cavity displacement α, followed by Wigner tomography with varying displacements β. A 280 ns waiting time after each measurement has been chosen to ensure that the readout cavity returns to the vacuum state. The qubit pulses have a Gaussian envelope truncated to $4\sigma = 8$ ns, and the displacement pulses on the storage cavity are 10 ns square pulses. The dashed enclosures represent the pulse sequence for a parity measurement.

To remove the cross-Kerr effect between the readout cavity and the storage cavity which skews the readout signal for large storage cavity displacements, and also to convert the readout voltage to parity, we followed the procedure in the supplementary material of ref. 14. The idea is to perform two parity measurements of different protocols ($R_{\hat{y}, \pi/2}$ or $R_{\hat{y}, -\pi/2}$ as the second $\pi/2$-pulse) for a vacuum state in the cavity. The difference between the two measurements corresponds to a parity value $P = 1$. We used this technique earlier in Extended Data Fig. 3 and here in the Wigner tomography in Extended Data Fig. 6c.

Extended Data Fig. 6d shows the error budgets for the Wigner tomography fidelity in Fig. 2. The fidelity is defined by the overlap between the measured Wigner state and that of an ideal cat state. We mainly consider qubit T_1, T_ϕ and photon jump processes. First of all, the qubit T_1 and T_ϕ processes and $|f\rangle$ states during Wigner tomography have been included in the parity calibration. After measurement M_1, we post-select only the $|g\rangle$ state of the qubit. Therefore, the system's initial state before the first parity measurement P_1 (note that $R_{\hat{y}, \pi/2}$ is the second $\pi/2$-pulse) is $|\psi_i\rangle = |\alpha\rangle|g\rangle$. The rotating frame has been fixed to be the one rotating at the cavity frequency ω_s when the qubit is in the $|g\rangle$ state. The qubit T_1 process between the two $\pi/2$-pulses in P_1 leads to $|\psi_f\rangle = |\alpha e^{i\chi_{qs}t}\rangle(|g\rangle + |e\rangle)/\sqrt{2}$, where t is the time of the T_1 jump happening. On average, this process gives 3% fidelity error. The qubit T_ϕ process between the two $\pi/2$-pulses in P_1 leads to $|\psi_f\rangle = [\mathcal{N}_+(|\alpha\rangle + |-\alpha\rangle)|g\rangle + \mathcal{N}_-(|\alpha\rangle - |-\alpha\rangle)|e\rangle]/\sqrt{2}$ right before measurement M_2, switching the entanglement relationship between the cat states and the qubit states and causing a full error. Here $\mathcal{N}_\pm = 1/\sqrt{2(1 \pm e^{-2|\alpha|^2})}$. Because in our experiment we mainly consider $|\alpha| = 2$, for reasons of simplicity, we use the approximation $\mathcal{N}_+ = \mathcal{N}_- = 1/\sqrt{2}$ in the large-$|\alpha|$ limit in what follows. The photon jump process between the two $\pi/2$-pulses in P_1 leads to $|\psi_f\rangle = [(|\alpha\rangle - e^{i\chi_{qs}t}|-\alpha\rangle)|g\rangle + (|\alpha\rangle + e^{i\chi_{qs}t}|-\alpha\rangle)|e\rangle]/2$, where t is the time of the photon jump event happening. On average, in this case the fidelity to the ideal even/odd cat states is 50%. If the photon jump process happens during measurement M_2 and the following waiting time, it switches between the even and odd cat states, thus causing a full error. If the photon jump process happens between the two $\pi/2$-pulses during the Wigner tomography, the results in the two parity protocols cancel out on average, leading to a full error as well. Finally, the qubit measurement inaccuracy due to the finite overlap between measurement histograms in P_1 plus the qubit transition up process during the waiting time right after the readout is about 1.3%. The sum of all sources of error limits the Wigner tomography fidelity $F = 84\%$ of the created cat states, consistent with the measurement. To create the even cat state, the measurement M_2 projects the qubit onto the $|e\rangle$ state. The extra qubit T_1 process after the projection lowers its creation fidelity by 280 ns/8 μs = 3.5%. This difference explains the imperfect cancellation of the fringes in Fig. 2c.

Similarly, the error budgets for the parity readout fidelity with $R_{\hat{y}, -\pi/2}$ as the second qubit pulse (Extended Data Fig. 6a) can also be estimated, as shown in Extended Data Fig. 6e. We again mainly consider qubit T_1, T_ϕ, $|f\rangle$ state and photon jump processes. We examine the case with the system initially in the state $|\psi_i\rangle = (|\alpha\rangle + |-\alpha\rangle)|g\rangle/\sqrt{2}$ (post-selected by the first five parity measurements), and consider the probability of not measuring the $|g\rangle$ state in the sixth parity measurement. In this case, the qubit T_1 process between the two $\pi/2$-pulses has a 50% chance of causing an error. The photon jump process between the two $\pi/2$-pulses leads to $|\psi_f\rangle = (|\alpha\rangle - |-\alpha\rangle)[(1 - e^{i\chi_{qs}t})|g\rangle - (1 + e^{i\chi_{qs}t})|e\rangle]/2\sqrt{2}$, where t is the time of the photon jump event happening, on average also giving a 50% chance of a wrong answer. The qubit T_ϕ process between the two $\pi/2$-pulses flips the qubit

state on the equator of the Bloch sphere and leads to $|\psi_f\rangle = (|\alpha\rangle + |-\alpha\rangle)|e\rangle/\sqrt{2}$, giving a full error in the final readout. Finally, the qubit measurement inaccuracy in the fifth parity measurement plus the qubit transition up process during the waiting time is 1.3%, as in the case of the Wigner tomography. There is an extra error coming from the $|f\rangle$ state between the two $\pi/2$-pulses in the sixth parity measurement, contributing about 0.5%. All the above sources of error add up to 7.7%, in good agreement with the 91.3% probability of faithfully measuring a positive correlation for an even cat state in Extended Data Fig. 6a. The lower fidelity for an odd cat state is because of the extra qubit T_1 process for the $|e\rangle$ state due to the negative correlation under the same parity readout protocol.

Quantum filter and correlated data. To mitigate the effects due to qubit decoherence, $|f\rangle$ states of the qubit (undesirable states that obscure the parity measurement) and other imperfections in the qubit readout in extracting the parity, we have applied a quantum filter[33,34] that best estimates the photon state parity. We note that the quantum filter is an integration of the quantum stochastic master equation and depends on the measured trajectory. Extended Data Fig. 7 shows the schematic of the quantum filter. This quantum filter at each point in time is realized in two steps: first, a new density matrix $\tilde{\rho}(C_{t+dt})$ is calculated from the best estimation $\rho(C_t)$ at the previous point, based only on the decoherence of the cavity; second, the density matrix $\tilde{\rho}(C_{t+dt})$ gets updated as the best estimation $\rho(C_{t+dt})$ according to Bayes' law, based on the newly acquired knowledge from the current parity measurement. This best estimated density matrix $\rho(C_{t+dt})$ is then used as the input for the next iteration. We have truncated the dimension of the density matrix to $N = 5\bar{n}$, which is large enough to cover all relevant number states. To initialize the density matrix after a displacement $D(\alpha)$, we have set $\rho(t=0) = (1 - n_{th})D(\alpha)|0\rangle\langle 0|D^\dagger(\alpha) + n_{th}D(\alpha)|1\rangle\langle 1|D^\dagger(\alpha)$, taking into account the background photon population in the limit $n_{th} \ll 1$.

At time t, the density matrix of the photon state is $\rho(C_t)$, which depends on all previous correlations. At $t + dt$, considering only the decoherence of the cavity, the expected density matrix from free evolution becomes $\tilde{\rho}(C_{t+dt}) = M_{down}\rho(C_t)M_{down}^\dagger + M_{up}\rho(C_t)M_{up}^\dagger + M_{no}\rho(C_t)M_{no}^\dagger$, where $M_{down} = \sqrt{\kappa_{down}dt}a$, $M_{up} = \sqrt{\kappa_{up}dt}a^\dagger$ and $M_{no} = I - \left(M_{down}^\dagger M_{down} + M_{up}^\dagger M_{up}\right)/2$ are the Kraus operators for photon loss, absorption of thermal photons and no jump events, respectively. We have $\kappa_{down} = (n_{th} + 1)\kappa$ and $\kappa_{up} = n_{th}\kappa$, and $\kappa = 1/\tau_{tot}$ is the energy decay rate in the cavity under repeated parity measurements. The additional information C_{t+dt} acquired from the parity measurement at $t + dt$ changes the quantum state according to

$$\rho(C_{t+dt}) = \begin{cases} P(even|C_{t+dt})\dfrac{\hat{P}_{even}\tilde{\rho}(C_{t+dt})\hat{P}_{even}}{Tr(\hat{P}_{even}\tilde{\rho}(C_{t+dt})\hat{P}_{even})} \\ \quad + P(odd|C_{t+dt})\dfrac{\hat{P}_{odd}\tilde{\rho}(C_{t+dt})\hat{P}_{odd}}{Tr(\hat{P}_{odd}\tilde{\rho}(C_{t+dt})\hat{P}_{odd})} & \text{if } C_{t+dt} \neq 0 \\ \tilde{\rho}(C_{t+dt}) & \text{if } C_{t+dt} = 0 \end{cases} \quad (1)$$

where $\hat{P}_{even}$ and $\hat{P}_{odd}$ are the projectors onto the even and odd manifolds, $\hat{P} = \hat{P}_{even} - \hat{P}_{odd} = e^{i\pi a^\dagger a}$ is the parity operator, and $P(even|C_{t+dt})$ and $P(odd|C_{t+dt})$ are the respective probabilities of being in the even- and odd-parity manifolds for a measured C_{t+dt}. To simplify the quantum filter, we assume that the event of the qubit jumping to the $|f\rangle$ states is independent of the cavity parity being even or odd. Hence, if the measured correlation is zero, the density matrix of the photon state is assigned to the one expected from free evolution. On the basis of Bayes' law, equation (1) becomes

$$\rho(C_{t+dt}) = \begin{cases} \dfrac{1}{P(C_{t+dt})}\left(P(C_{t+dt}|even)\hat{P}_{even}\tilde{\rho}(C_{t+dt})\hat{P}_{even}\right. \\ \quad \left. + P(C_{t+dt}|odd)\hat{P}_{odd}\tilde{\rho}(C_{t+dt})\hat{P}_{odd}\right) & \text{if } C_{t+dt} \neq 0 \\ \tilde{\rho}(C_{t+dt}) & \text{if } C_{t+dt} = 0 \end{cases}$$

where $P(C_{t+dt}) = P(C_{t+dt}|even)Tr\left[\hat{P}_{even}\tilde{\rho}(C_{t+dt})\hat{P}_{even}\right] + P(C_{t+dt}|odd)Tr\left[\hat{P}_{odd}\tilde{\rho}(C_{t+dt})\hat{P}_{odd}\right]$. The best parity estimation of the photon state is then

$$P(t+dt) = Tr\left[\rho(C_{t+dt})\hat{P}\right]$$

This formula has been used extensively in the main text to estimate the parity of the photon state.

To make a comparison with the best parity estimation on the basis of the above quantum filter, we also directly correlate the neighbouring parity measurements without any further processing. For zero correlation cases, because no information of the photon state parity is acquired, the best knowledge of parity at those points is just the last measured non-zero correlation. We assume the repeated parity measurement is a Markovian process. The ensemble-averaged parity dynamics obtained from the correlation under a repeated parity monitoring is then simply

$$\langle C_{cor}(t)\rangle = P(+1,t) - P(-1,t) \\ + P(0,t)\dfrac{P(+1,t) - P(-1,t)}{P(+1,t) + P(-1,t)} \quad (2)$$

where $P(+1,t)$, $P(-1,t)$ and $P(0,t)$ are the probabilities of measuring positive, negative and zero correlations at time t, respectively. The third term comes from the fact that the cases with zero correlation are assigned to cases previously measured to have non-zero correlation $+1$ or -1, whose probabilities are respectively $P(+1,t-dt)$ and $P(-1,t-dt)$. For small dt, $P(\pm 1, t-dt) \approx P(\pm 1, t)$.

The probabilities $P(+1,t)$, $P(-1,t)$ and $P(0,t)$ depend on both the measured parity readout property $P(\pm 1, 0|even/odd)$ and the even and odd parity evolutions, $P_e(t)$ and $P_o(t)$, of the photon state:

$$P(+1,t) = P(+1|even)P_e(t) + P(+1|odd)P_o(t)$$
$$P(-1,t) = P(-1|even)P_e(t) + P(-1|odd)P_o(t)$$
$$P(0,t) = P(0|even)P_e(t) + P(0|odd)P_o(t)$$

where $P_e(t) = \left(e^{-2|\alpha|^2 e^{-\kappa t}} + 1\right)/2$ and $P_o(t) = \left(1 - e^{-2|\alpha|^2 e^{-\kappa t}}\right)/2$.

With all the parameters in equation (2) known, $\langle C_{cor}(t)\rangle$ can be predicted. The agreement with the measured data is excellent, as shown in Extended Data Fig. 8a. This data set is the same as that shown in Fig. 4. Equation (2) even accurately predicts the offset in the averaged parity at $t = 0$, which comes from the asymmetric parity readout fidelities between the even and odd parities. The fact that the saturated parity value in the long time limit in Extended Data Fig. 8a is much lower than that in Fig. 4 is mainly a result of the qubit decoherence and the imperfections in the qubit readout. This large difference is additional proof of the effectiveness of the quantum filter. Extended Data Fig. 8b shows a direct comparison between the quantum filtered parity estimation and the direct correlation of qubit states between neighbouring parity measurements. The filtered data are clearly much smoother and can reject the brief changes in the correlated data, directly demonstrating the effectiveness of the quantum filter.

For a coherent state in a thermal bath, its parity dynamics is simply[21]

$$P = \frac{1}{1 + 2n_{th}}e^{-2|\alpha|^2 e^{-\kappa t}/(1+2n_{th})}$$

which has been used to fit the curves in Fig. 4.

Statistics of photon jumps. To test how faithfully our repeated parity measurement can track photon losses, we simply count the number of jumps extracted from the parity estimator during 500 μs repeated parity measurements. We have applied a Schmitt trigger to digitize the parity estimator to reject the unavoidable noise (spikes in the estimator) coming from qubit decoherence and erroneous parity readout. The two thresholds for the Schmitt trigger are chosen to be ± 0.9 for a large discrimination. Then the number of parity jumps is inferred from the number of transitions in the digital data after the Schmitt trigger.

Although our single parity readout fidelity is about 80% (Extended Data Fig. 6a), owing to the averaging effect of the quantum filter we actually can achieve nearly unity detection sensitivity of single-photon jump events. However, because of the finite bandwidth of the filter, if two photon jumps occur within the response time of the filter τ_f (defined as the time to make a transition between the two thresholds for the Schmitt trigger), our Schmitt trigger will not catch both jumps. Extended Data Fig. 8c shows the time response of the quantum filter applied to typical photon jump events. Green and cyan curves are fits of the parity estimator at the transition based on a hyperbolic tangent function, giving a transition time constant of less than 1 μs. We also find the response time of the filter to make a transition between ± 0.9 to be $\tau_f \approx 2$ μs. The probability of having a second photon jump within τ_f after the first jump is simply $P_{jump} = \dfrac{\bar{n}}{\tau_{tot}}\int_0^{\tau_f} e^{-t\bar{n}/\tau_{tot}}dt = 1 - e^{-\tau_f\bar{n}/\tau_{tot}}$. For $\bar{n} = 1$ and $\tau_{tot} = 49$ μs, the above probability is $P_{jump} = 4\%$, and $P_{jump} = 15\%$ for $\bar{n} = 4$, which is the probability of missing both jumps.

Extended Data Fig. 9 shows the histograms of the extracted number of jumps for an initial even or odd cat state obtained by post-selection. We note that the almost non-mixing distribution of even and odd numbers is trivial for the following reason. At the end of 500 μs repeated parity measurements, the cavity is already in a steady state with $n_{th} = 0.02$ photons, that is, 98% probability at vacuum (even parity) and 2% probability with one photon (odd parity). When the initial parity of the cat state, for example an even parity, is determined by post-selection, the number of jumps should have 98% probability of being even and only 2% probability of being odd, closely tied with the distribution of the final parity at $t = 500$ μs. A similar argument applies for an initially odd-parity cat state. The even–odd distributions in Extended Data Fig. 9 indeed show a 98-to-2% mixing, providing another way of determining n_{th}.

In reality, we have no way of knowing the true number of photon jumps for each parity measurement trajectory. The only way to test how faithfully our repeated parity measurement can track photon jumps is to see whether the distribution of jumps agrees with what we expect. Owing to the complication of background thermal excitation and the finite response time of the filter, to get an analytic solution is difficult. Instead, we perform a numerical Monte Carlo simulation to compare with the experiment. In the simulation, we use a coherent state as the initial state without distinguishing the parity. Each simulation trajectory is 500 μs long, and includes a transition probability of having a photon enter the cavity to change the photon number from n to $n + 1$ as a result of the background thermal excitation. In the simulation, we also neglect those who have neighbouring jumps within the response time τ_f of the quantum filter. Then for each trajectory we count the number of jumps and finally construct a histogram (black solid lines in Extended Data Fig. 9) of those numbers based on 100,000 trajectories. The good agreement between simulation and data demonstrates that the repeated parity measurement can track the error syndromes faithfully.

Quantifying parity-tracking performance. Our demonstrated parity-tracking protocol has several sources of infidelity that lead to a loss of the encoded information in the cat states, ultimately putting a bound on the improvement we would be able to achieve in an actual QEC protocol. This infidelity (Extended Data Fig. 6d) can be broken down into three categories: missed fast photon jumps (due to the limited bandwidth of the measurement), misinterpreted photon jumps (due to qubit T_ϕ and readout inaccuracy), and dephasing of the cat states due to the relaxation of the ancilla qubit during a parity measurement protocol (qubit T_1 process only, as explained later). Missing jumps certainly amounts to a complete loss of phase information. A distinction has to be made between the last two effects because misinterpreting photon jumps need not fully degrade our knowledge of the cavity state's parity at a given point in its trajectory and can be minimized by multiple repeated measurements (discussed more later). Recalling the Wigner tomogram in Fig. 2 aids in appreciating this point. Despite the 80% fidelity of a single parity measurement, after just three of them we can be very confident of the parity of our cavity state, because the probability of having three errant measurements is $(0.2)^3 = 0.8\%$. This is evident when inspecting the behaviour of the quantum filter in the single-shot traces; given three consecutive measurements that are the same, the filter converges to ± 1 with nearly 100% confidence. In the Wigner tomography, this would amount to knowing the value at the origin very well, but, given the cat state dephasing due to qubit T_1, not knowing the full contrast of the fringes and coherent state populations.

To realize the cat states as a quantum memory, the entire state must be preserved in order that an eventual decoding procedure[14] faithfully maps the information back onto some other component (for example a physical qubit). Given the long lifetime of our cavity, an 80% fidelity indicates that a large contribution to an incorrect parity measurement arises from two sources of qubit decoherence: T_1 and T_ϕ. The detrimental effects of T_1 decay are apparent when recalling the entanglement between the cavity and the ancilla qubit, where the cat state begins to acquire a phase at a rate $q\chi_{qs}$ that depends on the qubit state ($|g\rangle$ ($q = 0$), $|e\rangle$ ($q = 1$), $|f\rangle$ ($q = 2$)). Again, the rotating frame has been fixed to be the one rotating at the cavity frequency ω_s when the qubit is in the $|g\rangle$ state. Concretely, if qubit T_1 relaxation happens at t during the parity protocol waiting time π/χ_{qs}, the ideal cat state $(|\alpha\rangle \pm |-\alpha\rangle)/\sqrt{2}$ before the parity measurement will become $(|\alpha e^{i\chi_{qs}t}\rangle \pm |-\alpha e^{i\chi_{qs}t}\rangle)/\sqrt{2}$. Again, for reasons of simplicity, we have used the approximation $\mathcal{N}_+ = \mathcal{N}_- = 1/\sqrt{2}$ in the large-$|\alpha|$ limit for the rest of Methods. Similarly, during the time between parity measurements, the ideal cat state $(|\alpha\rangle \pm |-\alpha\rangle)/\sqrt{2}$ associated with different qubit states will rotate deterministically at a rate $q\chi_{qs}$. Should the qubit state change at a random time without our knowledge, the cat state will change its rotation rate accordingly, and the phase of the cat state will thus become completely random. In the Wigner tomography, this would manifest itself as a washing out of the cat state's features, where, unlike at the origin, successive measurements can only further reduce the fidelity. Thus, in a given single-shot record, qubit T_1 decay (and, indeed, excitations to higher qubit states) imparts an arbitrary phase on the cat states that would be impossible to recover from without some auxiliary correction protocol.

The contribution of qubit dephasing T_ϕ enters in a subtle, different way. Without loss of generality, let us assume the system is initially in state $(|\alpha\rangle + |-\alpha\rangle)|g\rangle/\sqrt{2}$. The first $\pi/2$-pulse in the parity measurement brings the system to $[(|\alpha\rangle + |-\alpha\rangle)|g\rangle + (|\alpha\rangle + |-\alpha\rangle)|e\rangle]/2$. The above state evolves to $[(|\alpha\rangle + |-\alpha\rangle)|g\rangle + (|\alpha e^{i\chi_{qs}t}\rangle + |-\alpha e^{i\chi_{qs}t}\rangle)|e\rangle]/2$, where at time t a random phase flip happens. Consequently, the system becomes $[(|\alpha\rangle + |-\alpha\rangle)|g\rangle - (|\alpha e^{i\chi_{qs}t}\rangle + |-\alpha e^{i\chi_{qs}t}\rangle)|e\rangle]/2$ and then keeps evolving in the same way, regardless of the sign change of the term associated with $|e\rangle$. At the end of the π/χ_{qs} evolution, the system becomes a product state again, $(|\alpha\rangle + |-\alpha\rangle)(|g\rangle - |e\rangle)/2$. The second $\pi/2$-pulse just takes the qubit to the other pole on the Bloch sphere, opposite to the case without the phase flip, indicating a 'fake' change in parity. Another way to understand the difference between qubit T_1 and T_ϕ errors is to recall the dispersive interaction between qubit and cavity as formulated in

the Hamiltonian in the main text: $(\omega_s - \chi_{qs}|e\rangle\langle e|)a^\dagger a$. Any change of the qubit energy results in a change of the cavity frequency, thereby dephasing the cavity state. A phase flip by contrast does not change the qubit energy, and the cavity phase therefore remains unaffected. Any extra phase acquired by the cavity state contingent on the final qubit state post-measurement would be completely deterministic (to within our ability to discriminate between $|e\rangle$ and $|g\rangle$). We note that the slow drift of the qubit frequency over time can also contribute to cavity dephasing. The cavity frequency change is of the order of $\chi_{qs}df/\Delta$, where df is the qubit frequency change and Δ is the qubit and cavity detuning. With a χ_{qs}/Δ of roughly 1/1,000, and a drift of qubit frequency of the order of 7 μs, we estimate this cavity dephasing to occur on time-scales of the order of 10 ms, much longer than any other timescale in our system. Therefore, this source of dephasing can be neglected. To summarize, qubit dephasing thus results in an incorrect parity measurement, but does not impart any erroneous phase onto the cavity state. Instead, by contributing to the 80% fidelity, the effect of T_ϕ necessitates more measurements to achieve a high degree of confidence in the estimation of the cavity state's parity. As increasing the number of measurements increases the likelihood of T_1 decay, qubit dephasing still facilitates an overall degradation of cat state fidelity.

The effective cavity decay rate, κ_{eff}, which predicts the gain one would expect to see in a parity-tracking protocol over the natural cavity photon jump rate $\bar{n}\kappa$, can be described by the sum of all dominant error rates including the three sources of infidelity described above plus a fourth to include readout errors. This κ_{eff} should also be a function of the number of consecutive parity measurements to make, N, and a waiting time τ_W between each one of these N-measurement 'packets' (Extended Data Fig. 10). Quantifying the parity-tracking performance as the foundation for a QEC protocol comes down to finding the optimal N and τ_W that give the lowest κ_{eff}, given the realistic experimental parameters at hand. Explicitly, we have

$$\kappa_{eff} = \left[\frac{(\bar{n}\kappa)^2 (N\tau_M + \tau_W)^2}{2} + N\epsilon_{T_1} + \mathcal{O}\left(\epsilon_{T_\phi}^m\right) + \mathcal{O}\left(\epsilon_M^m\right) \right] \frac{1}{N\tau_M + \tau_W}$$

where τ_M is the time it takes to perform one parity measurement, $\epsilon_{T_1} \approx \tau_M/T_1$, $\epsilon_{T_\phi} = \pi/\chi_{qs}T_\phi$ and ϵ_M are the losses of cat fidelity due to qubit T_1, T_ϕ and measurement inaccuracy, respectively (see section on Wigner tomography above), and $m = (N + 1)/2$.

Central to this idea is understanding that incorrect knowledge of the number of parity jumps is the first major contribution to cat state decoherence. Slow photon jump events can be tracked as demonstrated in the main text, and the corresponding phase errors can therefore be completely corrected. This means that single jumps can be tolerated. The term that goes like $(\bar{n}\kappa)^2$ determines the probability of instead having two or more jumps within the time window $N\tau_M + \tau_W$. Measurement infidelity in the form of qubit dephasing ϵ_{T_ϕ} and ϵ_M contributes to misleading the observer that a jump has occurred, which is equivalent to losing track of the number of jumps. These terms contribute to the κ_{eff} expression exponentially, and in a 'majority vote' fashion (assuming N is odd), which explains the choice of exponent in ϵ_{T_ϕ} and ϵ_M. Therefore, in principle the errors due to ϵ_{T_ϕ} and ϵ_M can drop out by either simple majority voting or the more elaborate quantum filtering. Collectively, these three terms correspond to cat state dephasing in the event of the incorrect tabulation of the number of jumps the cat state has undergone during its evolution.

The second major contribution to decoherence, ϵ_{T_1}, comes from cat state dephasing due to T_1. It takes only one qubit decay event within a string of N measurements to completely dephase a cat state, which is why ϵ_{T_1} grows linearly with N. Therefore, an optimal balance needs to be struck between gaining information about photon jumps and dephasing the cavity state by inquiring too frequently. To take an extreme case, we assume for a moment that there is no qubit decoherence and detection inaccuracy, so that ϵ_{T_1}, ϵ_{T_ϕ} and ϵ_M are all zero. In this case, each parity measurement would be perfect and the optimal parameters would thus be $\tau_W = 0$ and $N = 1$. In other words, without the threat of qubit-induced dephasing, the best approach would be to measure as frequently as possible. Conversely, with non-zero qubit decoherence, the longer the cavity lifetime, the more advantageous it would be to have a longer τ_W. Indeed, if $\kappa = 0$ then $\tau_W = \infty$. Should the parity tracking protocol become tolerant of qubit T_1 errors, ϵ_{T_1} would instead grow exponentially like ϵ_{T_ϕ}, because it would amount only to an overall loss of measurement fidelity.

Recalling the 83% fidelity of our Wigner tomography to an ideal cat state, the κ_{eff} derived here indicates that this fidelity does not decrease as 0.83^N, with N the number of parity measurements performed. The latter would be the case only if we were to encode a state, perform tomography on it with 83% fidelity, decode the state, whose fidelity to the initial one would then be reduced by 17%, and then repeat this procedure again and again. Of course this is not what we do. After the initial projection onto a cat state, we proceed with repeated parity measurements, and with each subsequent measurement we actually build up our confidence in the

state. At the same time, however, we pay the price of risking complete dephasing due to qubit decay. Thus, the number 0.83 reflects the amount of information we actually acquire when reconstructing the state through tomography rather than the degree to which the state has been corrupted. The actual corruption is related to how QND the measurement is regarding inducing extra photon loss and cavity state dephasing errors. In the main text, we show the former to be very high at 99.8% per measurement. The latter would be $1 - \epsilon_{T_1} \approx 94\%$.

Given our system's parameters, we can quantify what level of improvement we can achieve with the demonstrated parity-tracking protocol over a photon jump rate $\bar{n}\kappa$. As seen in Extended Data Fig. 6d, $\epsilon_{T_1} \approx 6\%$, $\epsilon_{T_\phi} \approx 3\%$ and $\epsilon_M \approx 1\%$ are all of the same order. Here we attribute a greater contribution of ϵ_{T_1} from that listed in Extended Data Fig. 6d because the sources of error there assumed that the final qubit state after each measurement was $|g\rangle$. However, without any post-selection of trajectories, the qubit could just as well end up in $|e\rangle$, enhancing the effect of qubit-induced dephasing. Therefore, the optimal N in our case is actually $N = 1$, notwithstanding the 80% fidelity of a single parity measurement. This can be understood by noting that for $N > 1$, the contribution of ϵ_{T_1} quickly begins to outweigh any qubit dephasing and measurement errors, leading to a suboptimal choice of parameters. Given that the contributions of these terms together sum to about a 10% error, we now have

$$\kappa_{\text{eff}} = \left[\frac{(\bar{n}\kappa)^2 (\tau_M + \tau_W)^2}{2} + 0.1 \right] \frac{1}{\tau_M + \tau_W}$$

The minimum κ_{eff} is achieved when the decay rates are equal:

$$(\tau_M + \tau_W)^2 = \frac{2(0.1)}{(\bar{n}\kappa)^2} \quad \Rightarrow \quad \kappa_{\text{eff}} = \bar{n}\kappa \sqrt{0.2}$$

The improvement over $\bar{n}\kappa$ is thus of the order of $\sqrt{0.2}$, which predicts an improvement in the effective cavity decay time by a factor of two over $1/\bar{n}\kappa$. The corresponding τ_W value is 4.6 μs. Given that τ_M is dominated in large part by the parity protocol waiting time π/χ_{qs}, a relevant benchmark for the overall performance becomes the product $\chi_{qs}T_1$. We emphasize that even for this system's modest coherence properties, an improvement by a factor of two would be impressive.

Indeed, if T_1 and T_ϕ approach 20 μs, the protected lifetime of the information would exceed 50 μs, the lifetime of a single-photon Fock state in the storage cavity.

The highly QND nature of the parity measurements at 99.8%, expanded on in the main text, was omitted from the analysis above owing to its minor contribution relative to all other sources of error. We consider this number to be one of the two figures of merit for the success of the parity-tracking protocol. If the very act of measuring parity were to induce photon jumps without our knowledge, the parity-tracking protocol itself would be flawed. This success, however, belies the degree to which we perturb the information stored in the cat states. Referring once more the Wigner tomograms, although we can confidently claim that we are QND as far as the point at the origin is concerned, given low qubit T_1 the same cannot be said of the rest of the information present in the Wigner tomogram at other points in I–Q space. In other words, the parity monitoring does not change photon number probabilities, but could change the relative phases between constituent Fock states. This second figure of merit, which can again be quantified as the contribution to cavity state dephasing due to qubit decay ($1 - \epsilon_{T_1} \approx 94\%$ per measurement) still clearly leaves much room for improvement. Nonetheless, although certainly presenting challenges, shortcomings arising from qubit performance and other higher-order mechanisms of dephasing not discussed here (such as self-Kerr of the cavity and cross-Kerr due to readout) do not seem insurmountable[35]. Addressing the issue of cavity state dephasing due to measurement is an important next step in improving the performance of this QEC scheme. We are confident that we can address the issue of qubit T_1 without substantially altering the parity-tracking protocol presented here, but we feel that this lies beyond the scope of this work.

31. Siddiqi, I. et al. Direct observation of dynamical bifurcation between two driven oscillation states of a Josephson junction. *Phys. Rev. Lett.* **94,** 027005 (2005).
32. Vijay, R., Devoret, M. H. & Siddiqi, I. The Josephson bifurcation amplifier. *Rev. Sci. Instrum.* **80,** 111101 (2009).
33. Belavkin, V. P. Quantum stochastic calculus and quantum nonlinear filtering. *J. Multivariate Anal.* **42,** 171–201 (1992).
34. Bouten, L., van Handel, R. & James, M. An introduction to quantum filtering. *SIAM J. Contr. Optim.* **46,** 2199–2241 (2007).
35. Mirrahimi, M. et al. Dynamically protected cat-qubits: a new paradigm for universal quantum computation. *New J. Phys.* **16,** 045014 (2014).
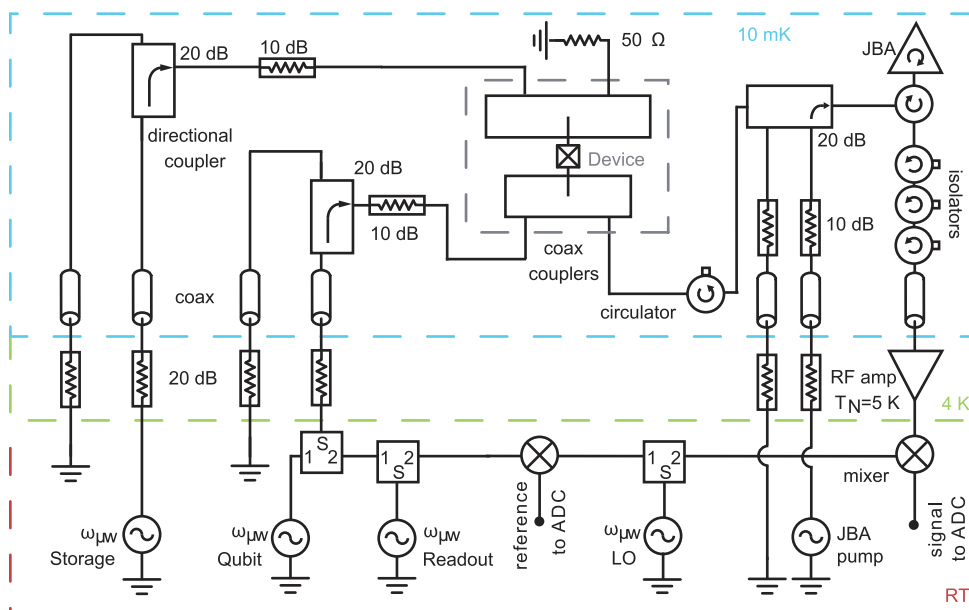
Extended Data Figure 1 | Schematic of the measurement set-up. We use two separate lines to drive the readout and the storage cavity. Qubit state manipulations are realized through the readout cavity input line. The readout cavity output signal is first amplified by a JBA operating in a double-pumped mode, and the reflected signal then goes through three isolators in series before being further amplified by a HEMT at 4 K. The amplified signal is finally down-converted to 50 MHz and then digitized by a fast 1 GS data-acquisition card.
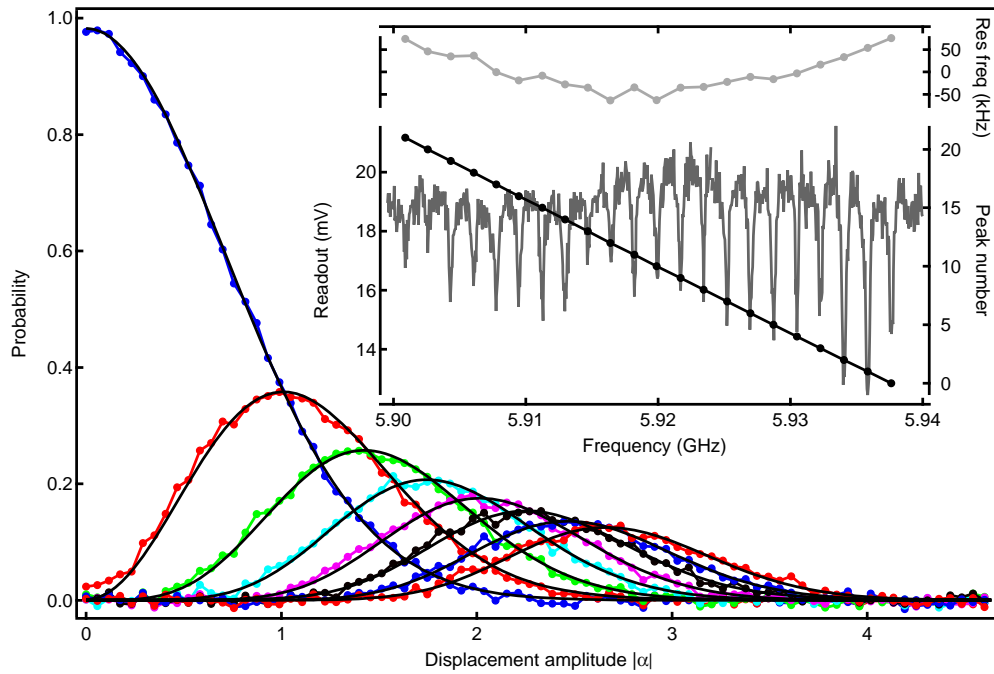
Extended Data Figure 2 | Poisson distribution of photon numbers in the cavity. Dotted colour lines are data for the first eight Fock states $n = 0, 1, 2, \ldots, 7$ as functions of displacement amplitude $|\alpha|$. The measurements are performed with a selective π-pulse on each number splitting peak, and the resulting signal amplitude should be proportional to the corresponding number population. These oscillation amplitudes have been normalized to probabilities such that the sum of the amplitudes corresponding to $n = 0$ and $n = 1$ equals unity. Dashed lines are theoretical curves with a Poisson distribution $P(|\alpha|, n) = |\alpha|^{2n} e^{-|\alpha|^2} / n!$, where the x axis has had a single scale factor adjusted to fit all these probabilities. The excellent agreement indicates good control of the coherent state in the cavity and also gives a good calibration of the cavity displacement amplitude. On the basis of the probability of $n = 1$ at $|\alpha| = 0$, we find a background photon population of $n_{th} = 0.02$ in the cavity. Inset bottom panel: spectroscopy (left axis) of the number splitting peaks of the qubit when populating different photon numbers in the cavity. Inset top panel: difference between peak positions and a linear fit. The curvature necessitates a second-order polynomial fit, resulting in a linear dispersive shift $\chi_{qs}/2\pi = 1.789 \pm 0.002$ MHz and a nonlinear dispersive shift $\chi'_{qs}/2\pi = 1.9 \pm 0.1$ kHz.
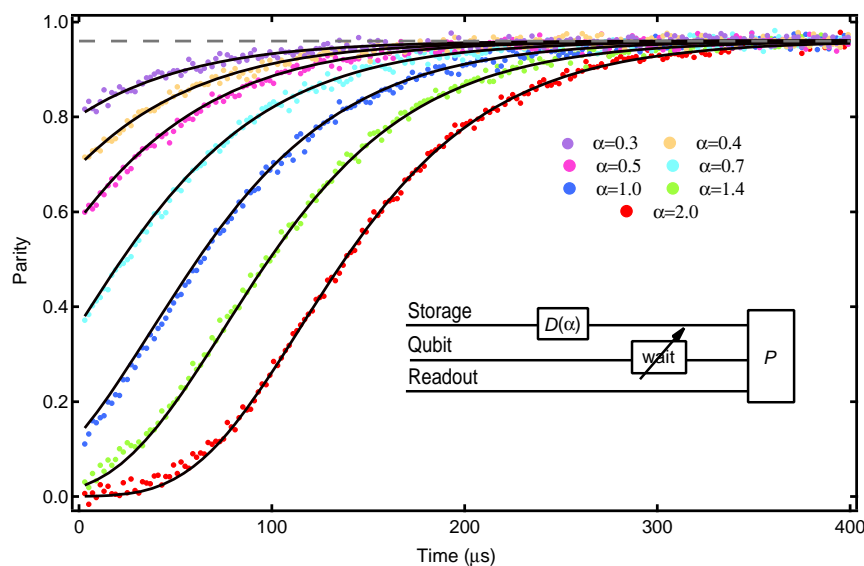
Extended Data Figure 3 | Ensemble-averaged free parity evolution of a coherent state. The measurement protocol is shown in the inset. The single parity measurement gives a readout voltage that has been converted to parity through thresholding. All measured evolution curves saturate at the same value in the long time limit. This saturation level has been forced to 0.96 (because $n_{th} = 0.02$), represented by the dashed horizontal line. The solid lines are global fits, giving a time constant of $\tau_0 = 55\,\mu s$.
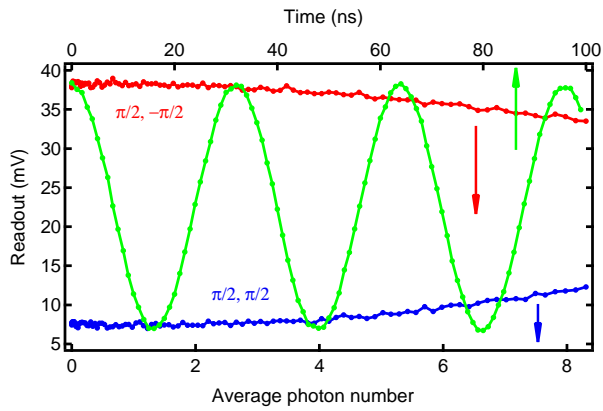
Extended Data Figure 4 | Effectiveness of the $R_{\hat{y}, \pm\pi/2}$ pulse. Blue and red data (bottom axis) are ensemble-averaged qubit readouts after consecutively (with no wait time) applying $(R_{\hat{y},\pi/2}, R_{\hat{y},\pi/2})$ and $(R_{\hat{y},\pi/2}, R_{\hat{y},-\pi/2})$, respectively, as functions of different $\bar{n}$ introduced into the cavity. The curvature for $\bar{n} > 4$ comes from the finite bandwidth of the pulses in the frequency domain. Green curve (top axis) is a time Rabi trace for an amplitude comparison with no initial cavity displacement.
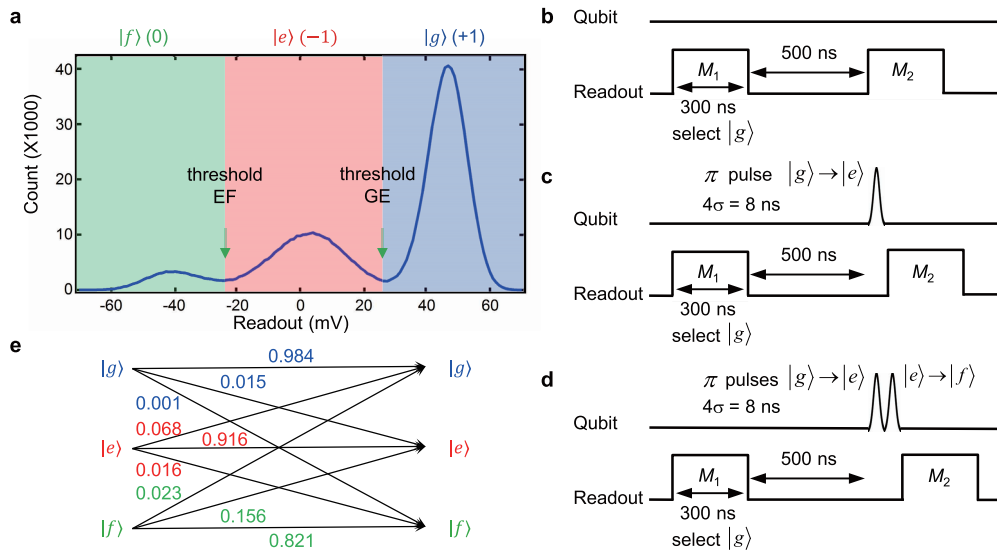
Extended Data Figure 5 | Qubit readout properties. a, Histogram of qubit readout for the parity protocol used in repeated single-shot traces in Fig. 3. The phase between the JBA readout and the pump has been adjusted such that $|g\rangle$, $|e\rangle$ and $|f\rangle$ states can be distinguished with optimal spacings. Thresholds between $|g\rangle$ and $|e\rangle$, and between $|e\rangle$ and $|f\rangle$, have been chosen to digitize the readout signal to $+1$, -1 and 0 for $|g\rangle$, $|e\rangle$ and $|f\rangle$, respectively. Note that we assign a zero to the $|f\rangle$ states to indicate a 'failed' measurement with no useful information about the parity. **b–d,** Illustrations of pulse sequences (not to scale) producing the readout error matrix with the storage cavity left in vacuum. The $|g\rangle$ state (**b**) is prepared through post-selection of an initial qubit measurement M_1, whereas $|e\rangle$ (**c**) and $|f\rangle$ (**d**) are prepared by properly pulsing the selected $|g\rangle$ state. A histogram of the second measurement, M_2, gives the qubit readout properties. **e,** Qubit readout properties for qubit initially in $|g\rangle$, $|e\rangle$ and $|f\rangle$, respectively.

d

source of errors	duration (μs)	coherence time (μs)	contribution to Winger tomography errors
qubit T_1 between two $\pi/2$ pulses in P_1	0.28	8.0	3.0%
qubit T_ϕ between two $\pi/2$ pulses in P_1	0.28	7.3	3.8%
photon jump between two $\pi/2$ pulses in P_1 ($\bar{n}=4$)	0.28	$50/\bar{n}=12.5$	1.1%
photon jump during $M2$ and waiting time in P_1 ($\bar{n}=4$)	0.58	$50/\bar{n}=12.5$	4.6%
photon jump in P_2 ($\bar{n}=4$)	0.28	$50/\bar{n}=12.5$	2.2%
qubit measurement inaccuracy in P_1	NA	NA	0.8%
qubit transition up after measuring $|g\rangle$ in P_1	0.28	56	0.5%

e

source of errors	duration (μs)	coherence time (μs)	contribution to parity measurement errors
qubit T_1 between two $\pi/2$ pulses	0.28	8.0	1.8%
qubit T_ϕ between two $\pi/2$ pulses	0.28	7.3	3.8%
photon jump between two $\pi/2$ pulses ($\bar{n}=1$)	0.28	50	0.3%
qubit measurement inaccuracy in P_5	NA	NA	0.8%
qubit transition up after measuring $|g\rangle$ in P_5	0.28	56	0.5%
qubit $|f\rangle$ state	0.28	56	0.5%

Extended Data Figure 6 | Parity readout properties and Wigner tomography. **a**, Parity readout property for given even and odd parity states for the protocol ($R_{\hat{y},-\pi/2}$ as the second qubit pulse) used in the single-shot traces in Fig. 3 ($\bar{n}=1$). **b**, Protocol to measure parity readout fidelity. An initial qubit measurement allows a post-selection of the $|g\rangle$ state of the qubit, followed by six consecutive parity measurements. The pulse sequence of each parity measurement is shown in P_1 in **c**. P($\pm1,0$|even/odd) are determined by post-selecting the cases with the first five consecutive identical parity results, which give the photon state parity with good confidence, and then constructing a histogram for the sixth parity measurement. **c**, Illustration of pulse sequence (not to scale) for producing the cat states and the Wigner tomography shown in

Fig. 2. The protocol starts with a post-selection of the $|g\rangle$ state of the qubit through an initial qubit measurement M_1. A parity measurement is performed immediately after a storage cavity displacement α, followed by Wigner tomography with varying displacements β. A 280 ns waiting time after each measurement has been chosen to ensure that the readout cavity is in the vacuum state. The qubit pulses have a Gaussian envelope truncated to $4\sigma = 8$ ns, and the displacement pulses on the storage cavity are 10 ns square pulses. The dashed enclosures represent the pulse sequences for parity measurement. **d**, Error budgets for Wigner tomography fidelity. **e**, Error budgets for the parity readout fidelities with $R_{\hat{y},-\pi/2}$ as the second qubit pulse.

$$P(\text{even}|C_{t+dt} \neq 0) \qquad \frac{\hat{P}_{\text{even}}\tilde{\rho}(C_{t+dt})\hat{P}_{\text{even}}}{\text{Tr}[\hat{P}_{\text{even}}\tilde{\rho}(C_{t+dt})\hat{P}_{\text{even}}]} \qquad \text{even}$$

$$\rho(C_t) \xrightarrow{P(\text{odd}|C_{t+dt} \neq 0)} \frac{\hat{P}_{\text{odd}}\tilde{\rho}(C_{t+dt})\hat{P}_{\text{odd}}}{\text{Tr}[\hat{P}_{\text{odd}}\tilde{\rho}(C_{t+dt})\hat{P}_{\text{odd}}]} \qquad \text{odd}$$

$$C_{t+dt}=0 \qquad \tilde{\rho}(C_{t+dt}) \qquad \text{no knowledge gained}$$
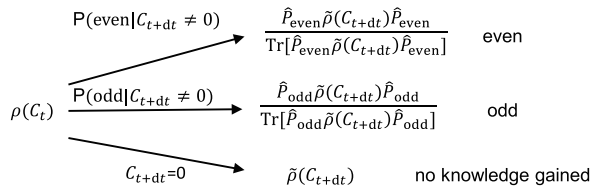
Extended Data Figure 7 | Schematic of the quantum filter. At time t, the density matrix of the photon state is $\rho(C_t)$, which depends on all previous correlations. At $t + dt$, only considering the decoherence of the cavity, the expected density matrix from free evolution becomes $\tilde{\rho}(C_{t+dt})$. The additional information C_{t+dt} acquired from the parity measurement at $t + dt$ changes the knowledge of the parity of the photon state according to equation (1).
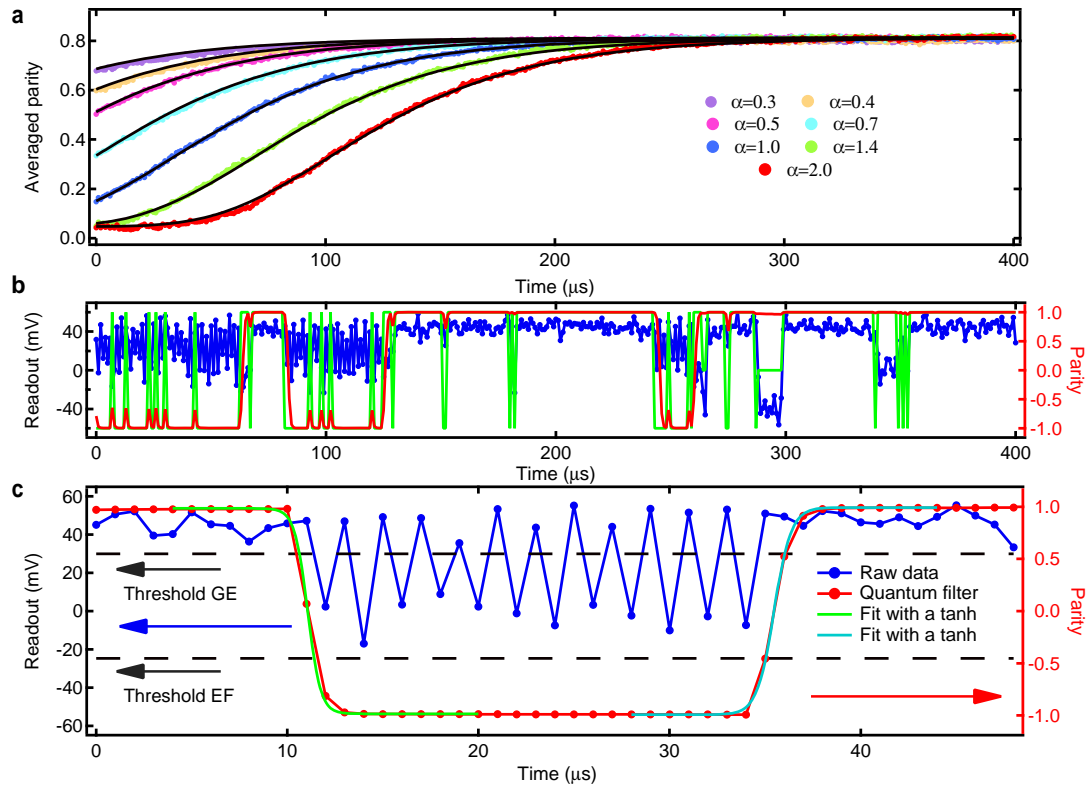
Extended Data Figure 8 | Effectiveness and response time of the quantum filter. a, Ensemble-averaged parity dynamics obtained directly from the correlation of qubit states between neighbouring parity measurements. The data set is the same as that shown in Fig. 4. Solid lines are predictions based on equation (2), in excellent agreement with the measured data. The offset of the averaged parity at $t = 0$ comes from the asymmetry between the parity readout fidelities of the even and odd states. The fact that the saturated parity value in the long time limit is much lower than that in Fig. 4 is additional proof of the effectiveness of the quantum filter. **b**, Effectiveness of the quantum filter. Blue (raw) and red (filtered) curves are the same as those shown in Fig. 3e.

The green curve is the direct correlation of qubit states between neighbouring parity measurements. The red curve is clearly much smoother and can reject the brief changes in the green curve. **c**, Response time of the quantum filter applied to typical photon jump events. The blue curve is the raw data from a repeated parity measurement. The red curve is the corresponding parity estimator based on the quantum filter. Green and cyan curves are fits to tanh functions of the parity estimator at the transitions down and up, respectively, giving a transition time constant of less than 1 μs. However, the response time of the filter to make a transition between −0.9 and +0.9 is $\tau_f \approx 2$ μs.
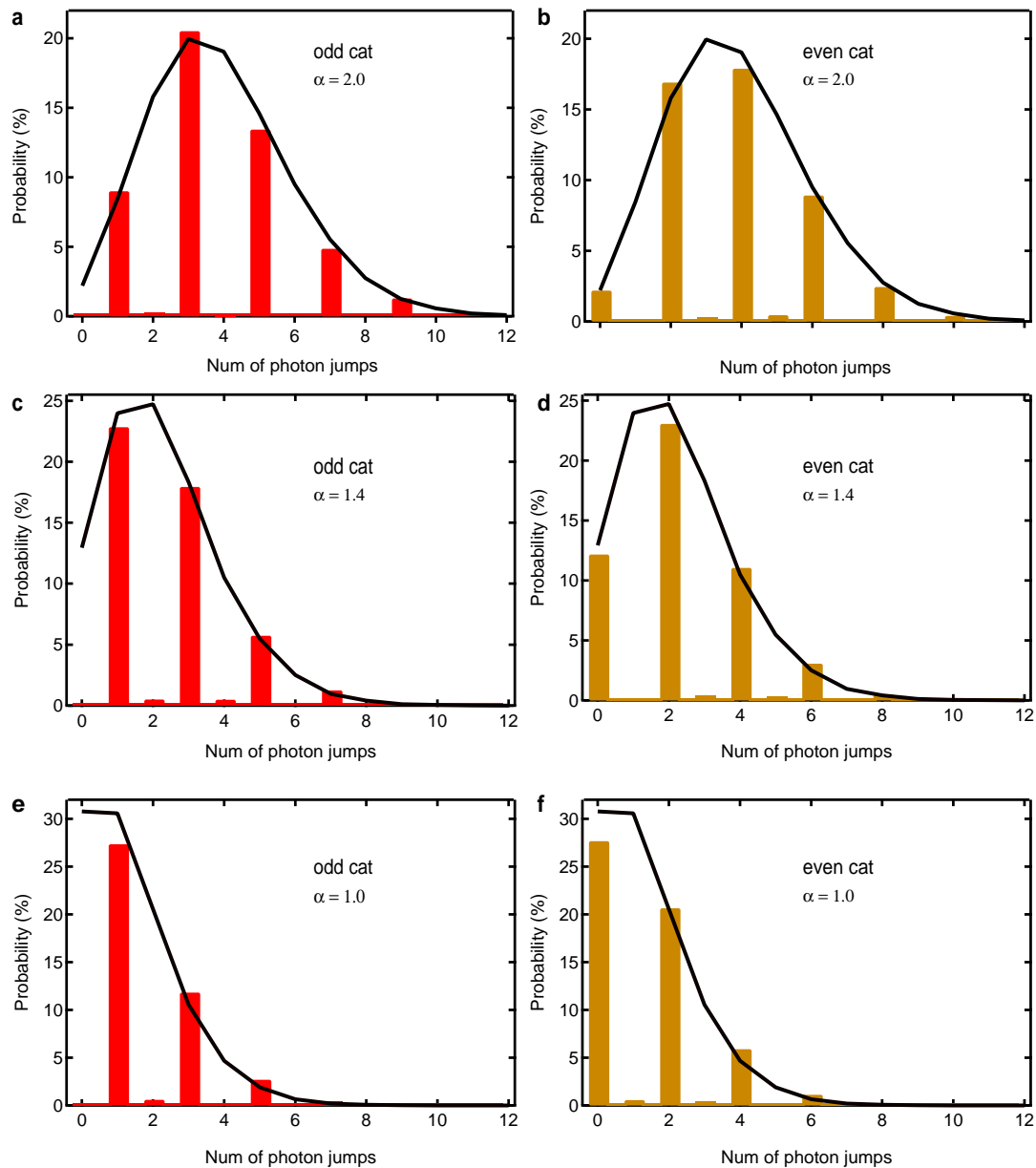
Extended Data Figure 9 | Histograms of the number of jumps extracted from the parity estimator during 500 μs repeated parity measurements for an initial even or odd cat state by post-selection. a, b, $|\alpha| = 2.0$; **c, d,** $|\alpha| = 1.4$; **e, f,** $|\alpha| = 1.0$. Solid lines are numerical simulations including the background thermal excitation and finite response time of the quantum filter. In the simulation, we use a coherent state as the initial state without distinguishing the parity. The good agreement between data and simulation demonstrates that the repeated parity measurement can track the error syndromes faithfully.
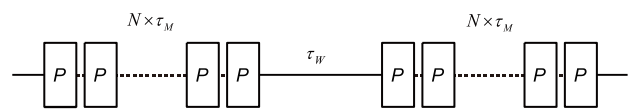
Extended Data Figure 10 | An optimized parity-tracking scheme would involve performing packets of N measurements, each lasting a time τ_M, followed by a waiting time of τ_W.

LETTER

Spin–transfer torque generated by a topological insulator

A. R. Mellnik[1], J. S. Lee[2], A. Richardella[2], J. L. Grab[1], P. J. Mintun[1], M. H. Fischer[1,3], A. Vaezi[1], A. Manchon[4], E.-A. Kim[1], N. Samarth[2] & D. C. Ralph[1,5]

Magnetic devices are a leading contender for the implementation of memory and logic technologies that are non-volatile, that can scale to high density and high speed, and that do not wear out. However, widespread application of magnetic memory and logic devices will require the development of efficient mechanisms for reorienting their magnetization using the least possible current and power[1]. There has been considerable recent progress in this effort; in particular, it has been discovered that spin–orbit interactions in heavy-metal/ferromagnet bilayers can produce strong current-driven torques on the magnetic layer[2–11], via the spin Hall effect[12,13] in the heavy metal or the Rashba–Edelstein effect[14,15] in the ferromagnet. In the search for materials to provide even more efficient spin–orbit-induced torques, some proposals[16–19] have suggested topological insulators[20,21], which possess a surface state in which the effects of spin–orbit coupling are maximal in the sense that an electron's spin orientation is fixed relative to its propagation direction. Here we report experiments showing that charge current flowing in-plane in a thin film of the topological insulator bismuth selenide (Bi_2Se_3) at room temperature can indeed exert a strong spin-transfer torque on an adjacent ferromagnetic permalloy ($Ni_{81}Fe_{19}$) thin film, with a direction consistent with that expected from the topological surface state. We find that the strength of the torque per unit charge current density in Bi_2Se_3 is greater than for any source of spin-transfer torque measured so far, even for non-ideal topological insulator films in which the surface states coexist with bulk conduction. Our data suggest that topological insulators could enable very efficient electrical manipulation of magnetic materials at room temperature, for memory and logic applications.

The proposed mechanism[16–19] that motivates our study of topological insulators as sources of current-induced spin-transfer torque is illustrated in Fig. 1a. When an in-plane current flows in the surface state of a topological insulator, more forward-going electron states are occupied than are backward-going states and, because of the helical locking of the relative orientations of the spin and the momentum in the surface state, this necessarily means that the flow of charge is accompanied by a non-equilibrium surface spin accumulation with the spin moment in the $-\hat{x}$ direction as depicted in Fig. 1a (a caret denotes a unit vector). If this spin accumulation couples to an adjacent magnetic film, the resulting flow of spin angular momentum will exert a spin-transfer torque on the magnet. We note that this mechanism is not related to any physics near the Dirac point of the topological insulator, which might be disrupted by coupling to a ferromagnet, and does not depend on having zero bulk conductivity within the topological insulator. This torque is related to the Rashba–Edelstein effect in non-topological materials[14]. However, the helical spin/momentum locking of the topological insulator surface state produces a different sign and a much larger magnitude of spin accumulation than is produced in non-topological materials (Methods).

Our samples are bilayers consisting of 8 nm of Bi_2Se_3 and 8 or 16 nm of permalloy, with an oxidized aluminium cap to prevent oxidation of the permalloy surface, patterned into strips 10–80 μm long and 2.5–24 μm wide. The Bi_2Se_3 is grown by molecular-beam epitaxy and the permalloy

is grown by sputtering (Methods). This sample geometry is actually not ideal for measuring torque due to current flow within the Bi_2Se_3, because the average resistivity of the Bi_2Se_3 is 25 or more times that of the metallic permalloy, such that the great majority of the current is shunted through the permalloy and does not contribute to the torque. Nevertheless, the torque from the Bi_2Se_3 is still strong enough to be measured accurately.

We determine the strength of current-induced torque by using a spin torque ferromagnetic resonance (ST-FMR) technique developed previously to measure the spin Hall torque from heavy metals[4,5]. Using the circuit shown in Fig. 1c, we apply a current of fixed microwave frequency and sweep an in-plane magnetic field through the ferromagnetic resonance condition. The oscillating current-induced torque causes the permalloy magnetization to precess, yielding resistance oscillations due to the anisotropic magnetoresistance (AMR) of the permalloy. We measure the resonance line shape using a direct voltage V_{mix}, which results from mixing between the applied alternating current and the oscillating resistance. The two vector components of the current-induced torque, in the $\hat{m} \times (\hat{x} \times \hat{m})$ ($\parallel$, in-plane) and $\hat{x} \times \hat{m}$ ($\perp$, perpendicular) directions (Fig. 1b), are respectively obtained from the amplitudes of the symmetric and antisymmetric components of the resonance line shape (Methods).
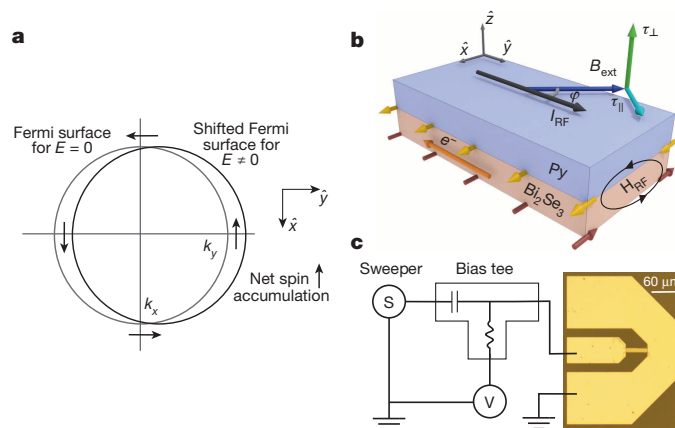
Figure 1 | The mechanism of current-induced spin accumulation in topological insulators and the sample geometry used in the measurement. a, Illustration of the mechanism by which an in-plane current in a topological insulator surface state generates a non-equilibrium surface spin accumulation, on account of locking between the spin direction and wavevector for electrons in the surface state. The arrows denote the directions of spin magnetic moments, which are opposite to the corresponding spin angular momenta because the g factor of the electron is negative. For simplicity, the spins in this cartoon are depicted in the sample plane, although some canting out of plane is expected. **b**, Schematic diagram of the layer structure and coordinate system. The yellow and red arrows denote spin moment directions. Py, permalloy. **c**, Depiction of the circuit used for the ST-FMR measurement and the sample contact geometry.

[1]Cornell University, Ithaca, New York 14853, USA. [2]Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. [3]Weizmann Institute of Science, Rehovot 76100, Israel. [4]King Abdullah University of Science and Technology, Physical Sciences and Engineering Division, Thuwal 23955-6900, Saudi Arabia. [5]Kavli Institute at Cornell, Ithaca, New York 14853, USA.

Figure 2a shows the results of the ST-FMR measurement for a 50 μm × 15 μm, 8 nm Bi_2Se_3/16 nm permalloy device with an in-plane magnetic field oriented at $\varphi = 45°$ relative to the current. The quality of the theoretical fit (to equation (1); see Methods) is excellent, and from the fit we determine that the oscillating torque per unit moment on the permalloy induced by the in-plane current has the components $\tau_\parallel = (2.7 \pm 0.3) \times 10^{-5}$ T and $\tau_\perp = (3.7 \pm 0.4) \times 10^{-5}$ T. The dependence on φ of both the symmetric and the antisymmetric components of V_{mix} is to good accuracy $V_{mix} \propto \cos(\varphi)^2 \sin(\varphi)$ (Fig. 2b). This is as expected for the ST-FMR signal (equation (1)), because the AMR resistance (R) satisfies $dR/d\varphi \propto \cos(\varphi)\sin(\varphi)$ and the spin-transfer torques arising from spin accumulation should give $\tau_\parallel, \tau_\perp \propto \cos(\varphi)$.

For comparisons with first-principles calculations, the quantities of primary interest are the effective spin current conductivities[22], $\sigma_{S,i} \equiv J_{S,i}/E = \tau_i M_s t_{mag}/[E\cos(\varphi)]$, where $J_{S,i}$ is the i component ($\parallel$ or $\perp$) of the spin current density absorbed by the ferromagnet for $\varphi = 0°$, E is the amplitude of the electric field, and $M_s t_{mag}$ is the product of the saturation magnetization and the thickness of the magnetic layer. For the measurement in Fig. 2a, the microwave field is $E = (0.8 \pm 0.1 \text{ V})/(50 \text{ μm})$ and $M_s t_{mag} = 14.2$ mA, as determined from vibrating-sample magnetometry on test films. Averaging over multiple devices with 8 nm Bi_2Se_3 and 16 nm permalloy, at frequencies of 6–10 GHz, we find that $\sigma_{S,\parallel} = (1.1 \pm 0.2) \times 10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$ and $\sigma_{S,\perp} = (1.4 \pm 0.2) \times 10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$. Performing the same measurement on 8 nm Bi_2Se_3/8 nm permalloy devices, we find $\sigma_{S,\parallel} = (2.0 \pm 0.4) \times 10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$ and $\sigma_{S,\perp} = (1.6 \pm 0.2) \times 10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$. These values are comparable to the spin current conductivities for the most efficient spin current sources known previously, namely heavy metals which generate current-induced spin currents by the spin Hall effect[4,5,23,24] (Table 1). Because the electrical conductivities of the heavy metals are much greater than that of Bi_2Se_3, this is a first indication that the strength of spin-transfer torque per unit current is greater in Bi_2Se_3 (see below). There is some potential uncertainty in our determination of $\sigma_{S,\parallel}$, in that pumping of spins from the precessing ferromagnet into the Bi_2Se_3, together with the inverse Edelstein effect, could produce an additional contribution to the symmetric ST-FMR signal that is not accounted for in our analysis. However, for this to affect our measurements appreciably would require a value of $|\sigma_{S,\parallel}|$ as large as or larger than that which we determine from the ST-FMR analysis (Methods). Therefore, the presence of a spin-pumping contribution would not change our central conclusion that Bi_2Se_3 provides very large values of $\sigma_{S,\parallel}$ and $\sigma_{S,\perp}$.

The sign we measure for $\sigma_{S,\parallel}$ is consistent with expectations for spin transfer from the current-induced spin accumulation in the Bi_2Se_3 topological insulator surface state with a chemical potential above the Dirac

Table 1 | Comparison of room-temperature $\sigma_{s,\parallel}$ and $\theta_{s,\parallel}$ for Bi_2Se_3 with other materials

Parameter	Bi_2Se_3 (this work)	Pt (ref. 4)	β-Ta (ref. 6)	Cu(Bi) (ref. 23)	β-W (ref. 24)
$\theta_\parallel$	2.0–3.5	0.08	0.15	0.24	0.3
$\sigma_{S,\parallel}$	1.1–2.0	3.4	0.8	—	1.8

$\theta_\parallel$ is dimensionless and the units for $\sigma_{s,\parallel}$ are $10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$.

point, that is, the sign corresponds to accumulation of spin angular momentum in the direction $-\hat{z} \times \hat{k}$, where $\hat{z}$ is the surface unit normal and $\hat{k}$ is the electron wavevector direction[25,26], or spin moment in the direction $\hat{z} \times \hat{k}$ (Fig. 1). The sign of $\sigma_{S,\perp}$ is the same as that of the torque due to an Oersted field, but the magnitude is much larger (see below).

An independent determination confirming the value of $\sigma_{S,\perp}$ can be obtained by measuring the current-induced shift in the ST-FMR resonant field (Extended Data Fig. 1 and Methods). Current-induced changes in the ST-FMR resonance width have been used previously as an alternative way of measuring $\sigma_{S,\parallel}$ (refs 2, 4), but we find that this is not possible in our samples because the resonance width is not a linear function of frequency (Extended Data Fig. 2 and Methods).

For applications, the figure of merit of primary interest is generally the 'spin torque ratio' $\theta_\parallel$, that is, the strength of the in-plane component of torque per unit applied charge current density J in the spin current source material ($\theta_\parallel \equiv (2e/\hbar)J_{S,\parallel}/J = (2e/\hbar)\sigma_{S,\parallel}/\sigma$, where σ is the charge conductivity in the spin current source material), because this quantity fundamentally determines the current needed for efficient magnetic manipulation[5,6,10,11]. Unlike $\sigma_{S,\parallel}$ and $\sigma_{S,\perp}$, which depend directly on the electric field, determining $\theta_\parallel$ requires knowing the average value of σ in Bi_2Se_3 when it is in contact with the permalloy. This is hard to determine because the addition of the permalloy to Bi_2Se_3 causes band bending in the Bi_2Se_3 that increases σ relative to its value in Bi_2Se_3 in isolation[26–28]. To mimic the band-bending effects of the permalloy, we have measured test samples of Bi_2Se_3 with an insulating Al_2O_x cap—made by depositing 2 nm of Al and oxidizing in a dry O_2–N_2 mixture—that also increase the chemical potential. For 8 nm of Bi_2Se_3 capped with Al_2O_x, we find at room temperature (22 °C) an average two-dimensional charge density of $9.4 \times 10^{13} \text{ cm}^{-2}$ by Hall measurements, and an average conductivity of $5.7 \times 10^4 \, \Omega^{-1} \text{m}^{-1}$. Angle-resolved photoemission spectroscopy has been used to measure the band bending of Bi_2Se_3 when coupled to various metals, including Ni and Rb (ref. 28). The band bending due to Rb was stronger than that due to Ni, and corresponded to a maximal doping of $5 \times 10^{13} \text{ cm}^{-2}$. We therefore make the rough estimate that our permalloy ($Ni_{81}Fe_{19}$) should increase the average conductivity of the Bi_2Se_3 by no more than the value we measure for the Al_2O_x cap. This yields a lower bound of $\theta_\parallel \gtrsim 3.5$ based on 8 nm permalloy devices (16 nm permalloy devices give a lower bound of $\theta_\parallel \gtrsim 2.0$). This is the largest spin torque ratio measured for any spin source material so far (Table 1). Control experiments on permalloy layers without Bi_2Se_3 and permalloy–Pt bilayers yield much smaller values of $\theta_\parallel$ and $\theta_\perp$ (Extended Data Fig. 3 and Methods).

The value that we determine for the out-of-plane spin current conductivity, $\sigma_{S,\perp} \approx 1.5 \times 10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$, is much larger than can be explained by the Oersted field. Given our estimates of σ for Bi_2Se_3 after band bending by the permalloy, the out-of-plane spin current conductivity that would be generated by the Oersted field alone is just $\sigma_{\perp,Oe} \approx 6.1 \times 10^3 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$ for the 8 nm permalloy samples and $\sigma_{\perp,Oe} \approx 1.2 \times 10^4 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$ for the 16 nm permalloy samples, smaller by factors of 23 and 13, respectively, than our measurements.

Our findings are in excellent agreement with a model of non-equilibrium spin accumulation near the Bi_2Se_3 surface and its diffusion into the permalloy layer (Extended Data Figs 4 and 5 and Methods). This model predicts correctly the signs of both $\sigma_{S,\parallel}$ and $\sigma_{S,\perp}$; torque due to the Rashba–Edelstein effect from non-topological interface states would give opposite signs. Furthermore, our model predicts, in agreement with the experiment, that the magnitude of $\sigma_{S,\perp}$ driven by the topological insulator surface state should have a magnitude comparable to that of $\sigma_{S,\parallel}$ and much larger than that produced by an Oersted field.
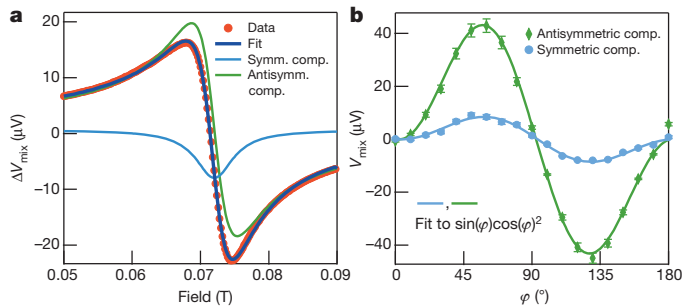

Figure 2 | ST-FMR measurements of the current-induced torque, with fits. **a**, Measured ST-FMR resonance at room temperature with microwave frequency $\omega/2\pi = 8$ GHz for an 8 nm Bi_2Se_3/16 nm permalloy sample of dimensions 50 μm × 15 μm. A fixed microwave power of 5 dBm is absorbed by the device ($I_{RF} = 7.7 \pm 1.1$ mA) and B_{ext} is oriented at an angle $\varphi = 45°$ from the current direction. The lines are fits to equation (1) (Methods) showing the symmetric and antisymmetric resonance components. **b**, Measured dependence on the magnetic field angle φ for the symmetric and antisymmetric resonance components for a different sample (8 nm Bi_2Se_3/16 nm permalloy device of dimensions 80 μm × 24 μm). Experimental conditions in **b** are as follows: $\omega/2\pi = 9$ GHz, power absorbed = 6.9 dBm. Error bars, 1 s.d.

Our findings have potential importance for technology, in that the spin torque ratio for Bi_2Se_3 at room temperature is larger than that for any previously measured spin current source material. However, as noted above, for practical applications the specific layer structure of our devices (topological insulator/metallic magnet) does not make good use of this high intrinsic efficiency because most of the applied current is shunted through the metallic magnet and does not contribute to spin current generation within the topological insulator. Applications will probably require coupling topological insulators to insulating (or high-resistivity) magnets so that the majority of the current will flow in the topological insulator. Using insulating magnets may also have the advantage of providing much less Gilbert damping than do metallic magnets[29], which can yield an additional reduction in the levels needed at present for manipulation via the spin torque anti-damping mechanism. Additionally, using an insulating magnet would allow electrostatic gating of bilayer devices, thus allowing the chemical potential to be tuned to eliminate any deleterious effects of bulk conduction. Our results therefore point towards a new strategy for implementing low-power, non-volatile magnetic memory and logic structures, using topological insulators as room-temperature sources of spin-transfer torque coupled to insulating magnetic layers, to achieve potential switching efficiencies higher than those produced by any other known mechanism.

Even stronger spin-transfer torques have been reported very recently at cryogenic temperatures in a topological insulator heterostructure[30].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Katine, J. A. & Fullerton, E. E. Device implications of spin transfer torques. *J. Magn. Magn. Mater.* **320,** 1217–1226 (2008).
2. Ando, K. *et al.* Electric manipulation of spin relaxation using the spin Hall effect. *Phys. Rev. Lett.* **101,** 036601 (2008).
3. Pi, U. H. *et al.* Tilting of the spin orientation induced by Rashba effect in ferromagnetic metal layer. *Appl. Phys. Lett.* **97,** 162507 (2010).
4. Liu, L. Q., Moriyama, T., Ralph, D. C. & Buhrman, R. A. Spin-torque ferromagnetic resonance induced by the spin Hall effect. *Phys. Rev. Lett.* **106,** 036601 (2011).
5. Miron, I. M. *et al.* Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection. *Nature* **476,** 189–193 (2011).
6. Liu, L. Q. *et al.* Spin torque switching with the giant spin Hall effect of tantalum. *Science* **336,** 555–558 (2012).
7. Kim, J. *et al.* Layer thickness dependence of the current-induced effective field vector in Ta|CoFeB|MgO. *Nature Mater.* **12,** 240–245 (2013).
8. Fan, X. *et al.* Observation of the nonlocal spin-orbital effective field. *Nature Commun.* **4,** 1799 (2013).
9. Garello, K. *et al.* Symmetry and magnitude of spin-orbit torques in ferromagnetic heterostructures. *Nature Nanotechnol.* **8,** 587–593 (2013).
10. Emori, S., Bauer, U., Ahn, S.-M., Martinez, E. & Beach, G. S. D. Current-driven dynamics of chiral ferromagnetic domain walls. *Nature Mater.* **12,** 611–616 (2013).
11. Ryu, K.-S., Thomas, L., Yang, S.-H. & Parkin, S. S. P. Chiral spin torque at magnetic domain walls. *Nature Nanotechnol.* **8,** 527–533 (2013).
12. Dyakonov, M. I. & Perel, V. I. Current-induced spin orientation of electrons in semiconductors. *Phys. Lett. A* **35,** 459–460 (1971).
13. Hirsch, J. E. Spin Hall effect. *Phys. Rev. Lett.* **83,** 1834–1837 (1999).
14. Edelstein, V. M. Spin polarization of conduction electrons induced by electric current in two-dimensional asymmetric electron systems. *Solid State Commun.* **73,** 233–235 (1990).
15. Chernyshov, A. *et al.* Evidence for reversible control of magnetization in a ferromagnetic material by means of spin–orbit magnetic field. *Nature Phys.* **5,** 656–659 (2009).
16. Burkov, A. A. & Hawthorn, D. G. Spin and charge transport on the surface of a topological insulator. *Phys. Rev. Lett.* **105,** 066802 (2010).
17. Culcer, D., Hwang, E. H., Stanescu, T. D. & Das Sarma, S. Two-dimensional surface charge transport in topological insulators. *Phys. Rev. B* **82,** 155457 (2010).
18. Pesin, D. & MacDonald, A. H. Spintronics and pseudospintronics in graphene and topological insulators. *Nature Mater.* **11,** 409–416 (2012).
19. Fischer, M. H., Vaezi, A., Manchon, A. & Kim, E.-A. Large spin torque in topological insulator/ferromagnetic metal bilayers. Preprint at http://arxiv.org/abs/1305.1328 (2013).
20. Fu, L., Kane, C. L. & Mele, E. J. Topological insulators in three dimensions. *Phys. Rev. Lett.* **98,** 106803 (2007).
21. Hasan, M. Z. & Kane, C. L. Topological insulators. *Rev. Mod. Phys.* **82,** 3045–3067 (2010).
22. Tanaka, T. *et al.* Intrinsic spin Hall effect and orbital Hall effect in 4d and 5d transition metals. *Phys. Rev. B* **77,** 165117 (2008).
23. Niimi, Y. *et al.* Giant spin Hall effect Induced by skew scattering from bismuth impurities inside thin film CuBi alloys. *Phys. Rev. Lett.* **109,** 156602 (2012).
24. Pai, C.-F. *et al.* Spin transfer torque devices utilizing the giant spin Hall effect of tungsten. *Appl. Phys. Lett.* **101,** 122404 (2012).
25. Yazyev, O. V., Moore, J. E. & Louie, S. G. Spin polarization and transport of surface states in the topological insulators Bi_2Se_3 and Bi_2Te_3 from first principles. *Phys. Rev. Lett.* **105,** 266806 (2010).
26. Bahramy, M. S. *et al.* Emergent quantum confinement at topological insulator surfaces. *Nature Commun.* **3,** 1159 (2012).
27. Wray, L. A. *et al.* A topological insulator surface under strong Coulomb, magnetic and disorder perturbations. *Nature Phys.* **7,** 32–37 (2011).
28. Valla, T., Pan, Z.-H., Gardner, D., Lee, Y. S. & Chu, S. Photoemission spectroscopy of magnetic and nonmagnetic impurities on the surface of the Bi_2Se_3 topological insulator. *Phys. Rev. Lett.* **108,** 117601 (2012).
29. Heinrich, B. *et al.* Spin pumping at the magnetic insulator (YIG)/normal metal (Au) interfaces. *Phys. Rev. Lett.* **107,** 066604 (2011).
30. Fan, Y. *et al.* Magnetization switching through giant spin-orbit torque in a magnetically doped topological insulator heterostructure. *Nature Mater.* **13,** 699–704 (2014).

Author Contributions A.R.M., J.S.L., A.R., N.S. and D.C.R. had the idea for and designed the experiments. A.R.M., J.L.G. and P.J.M. performed the sample fabrication, measurements and analysis. J.S.L., A.R. and N.S. developed the growth process for the Bi_2Se_3 layers. M.H.F., A.V., A.M. and E.-A.K. performed theoretical modelling. N.S. and D.C.R. provided oversight and advice. A.R.M. and D.C.R. wrote the manuscript and all authors contributed to its final version. Reprints and permissions information is available at www.nature.com/reprints

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.C.R. (dcr14@cornell.edu).

METHODS

Bilayer deposition and electrical properties. The Bi_2Se_3 thin films were grown by molecular-beam epitaxy on 2-inch C-axis oriented epi-ready sapphire wafers at a substrate temperature of approximately $300\,°C$ using thermal evaporation of high-purity (5N) elemental Bi and Se from Knudsen cells with a Se/Bi beam equivalent pressure ratio of approximately 14:1. The growth rate was 0.8 quintuple layers per minute. After growth, the films were capped with $\sim 4\,nm$ of Se to protect the surface. The Bi_2Se_3 film thickness was verified by X-ray reflection and the films were characterized by atomic force microscopy and high-resolution X-ray diffraction, yielding rocking curve widths less than $0.15°$. The samples were then transferred via air to a magnetron sputtering chamber with a base pressure of 2×10^{-9} Torr, where the Se capping layer was removed by heating the sample to $240\,°C$ for 1 hour. After cooling to room temperature, permalloy was sputtered on the bare Bi_2Se_3 followed by a 2 nm Al capping layer, which was subsequently oxidized in air. The average resistivity of the permalloy layer changes slightly with thickness owing to the roughness of the Bi_2Se_3, and is $71.7\,\mu\Omega$ cm for a nominally 8 nm layer and $45.9\,\mu\Omega$ cm for a nominally 16 nm layer. The average carrier density in the Se-capped Bi_2Se_3 was $2.64 \times 10^{19}\,cm^{-3}$ and n type, as determined by measurements of Hall effect (average areal density of $2.1 \times 10^{13}\,cm^{-2}$), such that the chemical potential lies above the Dirac point, at the edge of the conduction band. With the removal of the Se and the deposition of permalloy we anticipate a shift of the chemical potential even higher into the conduction band[26-28].

To make devices, we pattern the Bi_2Se_3/permalloy (Py) bilayers using optical lithography and ion milling, with electrical contacts made from 3 nm Ti/150 nm Pt in a symmetric geometry (Fig. 2b), so that when the samples are contacted using a ground–signal–ground high-frequency probe the currents travelling in the contacts do not produce a net Oersted field acting on the sample.

Analysis of ST-FMR measurements. We interpret the ST-FMR signals within a macrospin approximation for the magnetization direction $\hat{m}$ using the Landau–Lifschitz–Gilbert–Slonczewski equation of motion[31]:

$$\frac{d\hat{m}}{dt} = -\gamma \hat{m} \times (\vec{B}_{ext} - \mu_0 M_{eff} m_z \hat{z}) + \alpha \hat{m} \times \frac{d\hat{m}}{dt}$$
$$+ \gamma \tau_\parallel \frac{\hat{m} \times (\hat{x} \times \hat{m})}{|\hat{x} \times \hat{m}|} + \gamma \tau_\perp \frac{\hat{x} \times \hat{m}}{|\hat{x} \times \hat{m}|}$$

Here γ is the absolute value of the gyromagnetic ratio, $\vec{B}_{ext}$ is the applied magnetic field, $\mu_0 M_{eff}$ is the out-of-plane demagnetization field, α is the Gilbert damping constant, $\tau_\parallel$ is the in-plane component of the current-induced torque per unit moment (the symmetry analogous to a spin Hall torque), $\tau_\perp$ is the perpendicular component (analogous to the torque due to an Oersted or Rashba field), and $\hat{x}$ and $\hat{z}$ are defined as shown in Fig. 2a. We calculate that, near zero bias current and for small-angle precession, the ST-FMR mixing voltage has the form

$$V_{mix} = -\frac{I_{RF}\gamma}{4}\left(\frac{dR}{d\varphi}\right)\left[\tau_\parallel \frac{1}{\Delta}F_S(B_{ext},\omega) + \tau_\perp \frac{[1 + (\mu_0 M_{eff}/B_{ext})]^{1/2}}{\Delta}F_A(B_{ext},\omega)\right] \quad (1)$$

where I_{RF} is the total microwave current flowing through the device; $R(\varphi)$ is the anisotropic magnetoresistance as a function of the in-plane magnetization angle φ; $\Delta = \alpha\gamma(2B_{ext} + \mu_0 M_{eff})/2$ is the zero-current linewidth; $F_S(B_{ext},\omega) = (2\Delta)^2\omega^2/[(\omega^2 - \omega_0^2)^2 + (2\Delta)^2\omega^2]$ and $F_A(B_{ext},\omega) = [(\omega_0^2 - \omega^2)/2\omega_0\Delta]F_S(B_{ext},\omega)$ are approximately symmetric and, respectively, antisymmetric resonance line shapes as functions of B_{ext}, the magnitude of $\vec{B}_{ext}$; ω is the microwave frequency; and $\omega_0 \equiv \gamma\sqrt{B_{ext}(\mu_0 M_{eff} + B_{ext})}$ is the resonance frequency. These expressions are equivalent to those in ref. 4 except that we no longer make the approximations that $\Delta \ll B_{ext}$ and $B_{ext} \ll \mu_0 M_{eff}$. The two components of the current-induced torque can therefore be determined from the amplitudes of the symmetric and antisymmetric components of the resonance. Spin pumping, together with the inverse Edelstein effect, can provide an additional symmetric component to the resonance signal not accounted for here (see below).

For the ST-FMR analysis, we calibrate the anisotropic magnetoresistance $R(\varphi)$ for each device by rotating a 0.07 T magnetic field within the sample plane using a projective-field magnet (Extended Data Fig. 6). We use a network analyser to calibrate the transmission coefficients of our microwave circuit and the reflection coefficient of each sample. On the basis of this calibration, we determine I_{RF} and the strength of the electric field in the device for a given microwave power. The values we find from the ST-FMR measurements for $\sigma_{S,\parallel}$ and $\sigma_{S,\perp}$ are consistent for different applied microwave powers, demonstrating that the data correspond to the linear regime of small-angle precession, and they have no systematic variation with the length or width of the sample.

Theoretical modelling. We first consider the non-equilibrium spin accumulation of a topological surface state. The topological surface state of Bi_2Se_3 can be described by the Dirac Hamiltonian $\mathcal{H}_{\vec{k}}^D = v_F(\hat{z} \times \vec{\sigma}) \cdot \vec{k} - \mu_D$ where $\vec{\sigma}$ is the vector of Pauli

matrices acting in spin space, $\hat{z}$ is the unit vector in the z direction, and v_F and μ_D are constants. This leads to the characteristic linear dispersion and spin helicity around the Fermi surface of the topological insulator surface state. The velocity operator $\vec{v} = \partial_{\vec{k}}\mathcal{H}_{\vec{k}}^D$ is directly proportional to the spin operator $\vec{S} = (\hbar/2)\vec{\sigma}$ in the form $\vec{v} = 2v_F(\hat{z} \times \vec{S})/\hbar$. While the surface state has a vanishing equilibrium spin expectation value, any charge current flowing through the system leads to a non-zero spin accumulation independent of the microscopic details of the transport. A current density $j_y^D = en\langle v\rangle_{neq}$ in the y direction (where e is the electron's charge and n is the electron density) produces a spin density in the x direction:

$$\langle S_x\rangle_{neq}^D = \frac{\hbar}{2ev_F}j_y^D \quad (2)$$

Next we consider the spin accumulation due to the existence of additional non-topological surface states. The Bi_2Se_3/ferromagnet interface is expected to host not only a topological Dirac state but also a Rashba-split two-dimensional electron gas (2DEG) with a dispersion $\mathcal{H}_{\vec{k}}^R = \frac{\vec{k}^2}{2m_R} + \alpha_R(\hat{z} \times \vec{\sigma}) \cdot \vec{k} - \mu_R$ (ref. 32). Extended Data Fig. 4a shows schematically the combined band structure for the case with band offset parameters $\mu_D > \mu_R > 0$. Both v_F and α_R are positive, so that an alternating spin structure with a clockwise spin angular momentum on the topological insulator surface state (corresponding to an anticlockwise spin moment) is obtained[32] (Extended Data Fig. 4b).

The Rashba-split surface state leads to an additional spin accumulation $\langle S_x\rangle_{neq}^R = -\frac{\hbar}{2e}\frac{m_R\alpha_R j_y^R}{2E_F}$ (refs 33, 34), where E_F is the Fermi energy of the Rashba state. The spin accumulation of a Rashba-split surface state is partially compensated because the two sub-bands that are split by the spin–orbit coupling contribute to the spin accumulation with opposite signs, with the result that the spin accumulation from a Rashba-split surface state is smaller by a factor of $2\Delta k_F/k_F$ (Δk_F is the splitting of the bands at E_F) by comparison with the Dirac state.

The total spin accumulation is

$$\langle S_x\rangle_{neq} = \frac{\hbar}{2e}\left(\frac{j_y^D}{v_F} - \frac{m_R\alpha_R j_y^R}{2E_F}\right)$$

As long as v_F and α_R have the same sign, the non-equilibrium spin accumulation due to the Rashba-split 2DEG and that due to the helical Dirac state are opposite. Judging from the sign of the torque we measure experimentally, we can identify the topological surface state as the dominant source for the generation of torque in our experiments.

Now we turn to analysing the magnitude and direction of the spin-transfer torque that is generated by the non-equilibrium spin accumulation. In the following, we model the torque generation in permalloy through diffusion of the spin accumulation from the Bi_2Se_3 surface (equation (2)) into the permalloy. The diffusion (in the z direction) leads to a steady-state (itinerant) spin density determined by[35]

$$0 = -\vec{\nabla} \cdot \vec{\mathcal{J}}_i - \frac{1}{\tau_J}(\vec{S} \times \hat{m})_i - \frac{1}{\tau_\phi}[\hat{m} \times (\vec{S} \times \hat{m})]_i - \frac{S_i}{\tau_{sf}} \quad (3)$$

where the spin current (for the ith spin component) is given by $\vec{\mathcal{J}}_i = -\mathcal{D}\vec{\nabla}S_i$ with $\mathcal{D}$ the diffusion coefficient. The second term in equation (3) describes the precession of the spins around the moment of the permalloy, with τ_J the spin precession time. The third term captures the relaxation of the spin component perpendicular to the permalloy magnetization direction $\hat{m}$, with τ_ϕ the spin decoherence time. The last term describes spin diffusion with a timescale τ_{sf}.

We solve equation (3) subject to the requirement that there is no spin current through the outer boundary of the permalloy, that is, $\vec{\mathcal{J}}_i(d) = 0$, where d is the thickness of the permalloy layer. For the Bi_2Se_3/Py interface, we assume that, owing to the exchange interaction, the itinerant spins of the permalloy at the interface align with the spin density of the topological insulator interface, that is, $\vec{S}(0) = \chi\langle\vec{S}\rangle_{neq}$ with χ of order 1 (we will set $\chi = 1$ in the following). Note that this choice of the boundary condition for the diffusion equation is crucial. For the contrasting case of a pure spin-Hall-effect geometry, the torque would be due to a spin current injected into the ferromagnet, and the correct boundary condition would thus be a non-zero spin current at the interface, that is, at the interface $\langle\vec{S}\rangle_{neq} \approx 0$ and $\vec{\mathcal{J}}_i(0) \neq 0$. For realistic parameters, $\tau_{sf} \gg \tau_\phi, \tau_J$, this results in a spin Hall torque that is almost completely in-plane.

For the boundary conditions corresponding to non-equilibrium spin accumulation at the Bi_2Se_3/Py interface, the solution for the spin distribution in the z direction is

$$\hat{S}(z) = S_\parallel(z) + iS_\perp(z) = S_0\frac{\cosh[k(z-d)]}{\cosh(kd)} \quad (4)$$

with $k = (\lambda_\parallel^{-2} - i\lambda_J^{-2})^{1/2}$ and $\lambda_\parallel^{-2} = \lambda_{sf}^{-2} + \lambda_\phi^{-2}$, where $\lambda_J^2 = \mathcal{D}\tau_J$, $\lambda_{sf}^2 = \mathcal{D}\tau_{sf}$ and $\lambda_\phi^2 = \mathcal{D}\tau_\phi$. $S_\parallel(z)$ is the in-plane spin density and S_0 is the initial spin density (at $d = 0$),

both of which are perpendicular to $\hat{m}$. The torque on the permalloy moments is given by the spatial change of the spin current compensated by the spin relaxation

$$\hat{T} = \int_0^d dz \left[-\partial_z \hat{\mathcal{J}}(z) - \frac{1}{\tau_{\text{sf}}} \hat{S}(z) \right]$$

where we again use the short forms $\hat{T} = T_\parallel + iT_\perp$ and $\hat{\mathcal{J}} = \mathcal{J}_\parallel + i\mathcal{J}_\perp$. Given the spin distribution in the z direction of equation (4) and using $d \to \infty$, we find

$$\hat{T} = S_0 \frac{\mathcal{D}}{k} \left(\frac{1}{\lambda_\phi^2} - \frac{i}{\lambda_J^2} \right)$$

For $\varphi = 45°$ (Fig. 2a), the spin polarization perpendicular to the magnetization of the permalloy is $1/\sqrt{2}$ of the total spin accumulation at the interface (equation (2)), and we find

$$\hat{T} = -\frac{\hbar}{2} \frac{\mathcal{D}}{k} \left(\frac{1}{\lambda_\phi^2} - \frac{i}{\lambda_J^2} \right) \frac{\sqrt{2}}{2} \frac{j_y}{ev_F}$$

In analogy to the spin Hall angle $\theta_{\text{SH}} = (2eJ_S)/\hbar J$, the spin torque ratio for the Bi$_2$Se$_3$/ferromagnet interface is

$$\hat{\theta} = \frac{\hat{T}}{j_y} \frac{2e}{\hbar} = -\frac{\sqrt{2}}{2} \frac{\mathcal{D}}{v_F k} \left(\frac{1}{\lambda_\phi^2} - \frac{i}{\lambda_J^2} \right)$$

Using realistic values for the parameters ($\lambda_J = \lambda_\phi = 1$ nm, $\lambda_{\text{sf}} = 5$ nm (ref. 36), $v_F = 5 \times 10^5$ m s^{-1}) and a diffusion coefficient of $\mathcal{D} \approx 10$ cm^2 s^{-1}, we find in-plane and out-of-plane spin torque ratios of order one. Extended Data Fig. 5 shows the relative value of out-of-plane to in-plane spin torque ratios $\theta_\perp/\theta_\parallel$ as a function of λ_J/λ_ϕ to illustrate that for $\lambda_J \approx \lambda_\phi$ the two torque components are roughly the same size. Our experimental finding of a large spin torque ratio for both in-plane and out-of-plane directions is thus consistently described as arising from the topological surface state of Bi$_2$Se$_3$. We note that if the magnetization is spatially inhomogeneous through the thickness of the permalloy film, this could alter the form of equation (4) and provide an additional mechanism to change the ratio of the out-of-plane and in-plane torques.

Control experiments on single-layer permalloy samples and Pt/Py bilayers. As a first control experiment, we compared the ST-FMR signals measured for Bi$_2$Se$_3$/Py bilayers to single-layer permalloy films deposited on sapphire (Extended Data Fig. 3a). As has already been reported[4], the ST-FMR signal for single-layer permalloy samples is not zero. For the single-layer permalloy, we observe a small antisymmetric resonance that may be the result of non-uniform current density within the permalloy. However, the signal from the single-layer permalloy is small enough to represent only a small correction to the Bi$_2$Se$_3$/Py results.

We also performed ST-FMR measurements on Pt/Py bilayers consisting of 6 nm of sputtered Pt and 16 nm of sputtered permalloy. The results for one such device are shown in Extended Data Fig. 3b. The resistivity of the Pt was 22 μΩ cm and $M_s t_{\text{mag}} = 14.5$ mA for the sputtered permalloy films. Averaging over multiple devices and frequencies of 6–10 GHz, we find $\sigma_{\text{S},\parallel}^{\text{Pt}} = (4.4 \pm 1.1) \times 10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$ and the spin torque ratio $\theta_\parallel^{\text{Pt}} \approx 0.09 \pm 0.03$, consistent with our group's previously reported values[4]. We also find that $\sigma_{\text{S},\perp}^{\text{Pt}} = (8.0 \pm 1.8) \times 10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$, similar to the expected contribution from the Oersted field, $\sigma_{\perp,\text{Oe}}^{\text{Pt}} \approx 1.0 \times 10^6 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$.

Measurements of $\sigma_{\text{s},\parallel}$ based on the current dependence of the ST-FMR. A direct current produces a small shift in the centre field of the ST-FMR signal at fixed frequency (Extended Data Fig. 1). This can be explained as due to a current-induced out-of-plane d.c. spin torque $\tau_{\perp,d.c.}$, which to a good approximation alters the ST-FMR signal in the same way as a shift in the applied magnetic field, $\vec{B}_{\text{ext}} \to \vec{B}_{\text{ext}} - \tau_{\perp,d.c.} \tan(\varphi)$. By this method we estimate $\sigma_{\text{S},\perp} = (2.2 \pm 0.6) \times 10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$ in 8 nm Bi$_2$Se$_3$/16 nm permalloy devices, which is consistent, within the experimental uncertainty, with our ST-FMR measurement in these devices, $\sigma_{\text{S},\perp} = (1.4 \pm 0.2) \times 10^5 \hbar/2e \, \Omega^{-1} \text{m}^{-1}$.

Anomalous behaviour of the ferromagnetic resonance linewidths. In heavy-metal/ferromagnet samples, the in-plane current-induced torque due to the spin Hall effect can be determined accurately using measurements of the current dependence of the FMR linewidth, because the in-plane torque can alter the effective Gilbert damping[2,4]. We attempted to perform this analysis for our Bi$_2$Se$_3$/Py samples, but encountered the difficulty that the linewidth is not simple a linear function of the microwave frequency (Extended Data Fig. 2), such that we could not extract a meaningful measurement of the Gilbert damping parameter. We do not know the reason for this unusual behaviour. One possible speculation (that we will pursue in future measurements) is that the interaction with the Bi$_2$Se$_3$ might cause spatially non-uniform magnetic configurations in the permalloy film that evolve as a function of applied magnetic field in the range corresponding to the data in Extended Data Fig. 2.

Consequences of spin pumping and the inverse Edelstein effect. In analogy with calculations of spin pumping and the inverse spin Hall effect[37,38], spin pumping, together with the inverse Edelstein effect, should contribute to the ST-FMR measurement a signal with a symmetric line shape and a magnitude

$$V_{\text{sp}} = \theta_\parallel \frac{ew\lambda_{\text{BiSe}} R}{2\pi} \tanh\left(\frac{t_{\text{BiSe}}}{2\lambda_{\text{BiSe}}} \right) \text{Re}(g_{\uparrow\downarrow}^{\text{eff}}) \left\langle \left(\vec{m} \times \dot{\vec{m}} \right)_x \right\rangle$$

Here w is the sample width, λ_{BiSe} is the length scale for the thickness of the Bi$_2$Se$_3$ involved in spin–charge conversion, R is the sample resistance, φ is the magnetic field angle, t_{BiSe} is the thickness of the Bi$_2$Se$_3$ layer, and $\text{Re}(g_{\uparrow\downarrow}^{\text{eff}})$ is the real part of the effective spin mixing conductance. The sign of this contribution is opposite to that of the contribution to the symmetric part of the ST-FMR signal from the in-plane component of spin torque and AMR. The dependence on the angle of the applied magnetic field is expected to be the same as shown in Fig. 2b, with the result that the spin pumping and AMR contributions cannot be distinguished on this basis.

To estimate a rough upper bound on $|V_{\text{sp}}|$, we note that the value of $\lambda_{\text{BiSe}} \tanh(t_{\text{BiSe}}/2\lambda_{\text{BiSe}})$ is always less than $t_{\text{BiSe}}/2$, and so we will use this as an upper limit. For $\text{Re}(g_{\uparrow\downarrow}^{\text{eff}})$, we will use the value determined in Py/Pt samples[39], $\text{Re}(g_{\uparrow\downarrow}^{\text{eff}}) \approx 2 \times 10^{19}$ m^{-2}, although this might be an overestimate because the conductivity of Pt is much larger than that of Bi$_2$Se$_3$. For precession in a magnetic thin film with an out-of-plane demagnetization field $\mu_0 M_{\text{eff}}$, we calculate

$$\left\langle \left(\vec{m} \times \dot{\vec{m}} \right)_x \right\rangle = \omega\phi_{\text{p}}^2 \sin(\varphi) \sqrt{\frac{B_{\text{ext}}}{B_{\text{ext}} + \mu_0 M_{\text{eff}}}}$$

where ϕ_{p} is the maximum precession angle (in the sample plane). For the experimental conditions corresponding to the data in Fig. 2a, we find that $\phi_{\text{p}} = \frac{1}{dR/d\varphi} \frac{2}{I_{\text{RF}}} \sqrt{(V_{\text{mix}}^S)^2 + (V_{\text{mix}}^A)^2} \approx 0.0095$, where V_{mix}^S and V_{mix}^A are the coefficients of F_S and F_A determined by fitting the V_{mix} signal to equation (1). With these assumptions and using $\theta_\parallel = 2.0$–3.5, we find an upper bound for the spin-pumping voltage of $V_{\text{sp}} \approx 5$–9 μV, which is of the same order as the symmetric ST-FMR signal shown in Fig. 2a. Therefore, depending on the actual values of λ_{BiSe} and $\text{Re}(g_{\uparrow\downarrow}^{\text{eff}})$, the spin-pumping voltage might represent an appreciable correction to the ST-FMR analysis of the symmetric part of the ST-FMR resonance. However, to do so would in any case require a spin torque ratio $|\theta_\parallel|$ much greater than 1, so that, regardless of whether or not V_{sp} is large enough to affect the measurement, our conclusion that Bi$_2$Se$_3$ provides very large values of $\theta_\parallel$ and $\sigma_{\text{S},\parallel}$ remains unchanged. Because spin pumping and the inverse Edelstein effect contribute only to the symmetric line shape, any presence of these effects should not alter our interpretation of the antisymmetric part of the ST-FMR signals and the large value we determine for $\sigma_{\text{S},\perp}$.

Other spin torque effects involving topological insulators. Other types of spin torque effect involving topological insulators have also been proposed[40–44], but we do not believe that these are pertinent to our sample geometry and doping level.

31. Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159,** L1–L7 (1996).
32. King, P. D. C. *et al.* Large tunable Rashba spin splitting of a two-dimensional electron gas in Bi$_2$Se$_3$. *Phys. Rev. Lett.* **107,** 096802 (2011).
33. Manchon, A. & Zhang, S. Theory of nonequilibrium intrinsic spin torque in a single nanomagnet. *Phys. Rev. B* **78,** 212405 (2008).
34. Pesin, D. A. & MacDonald, A. H. Quantum kinetic theory of current-induced torques in Rashba ferromagnets. *Phys. Rev. B* **86,** 014416 (2012).
35. Manchon, A., Matsumoto, R., Jaffres, H. & Grollier, J. Spin transfer torque with spin diffusion in magnetic tunnel junctions. *Phys. Rev. B* **86,** 060404 (2012).
36. Bass, J. & Pratt, W. P. Spin-diffusion lengths in metals and alloys, and spin-flipping at metal/metal interfaces: an experimentalist's critical review. *J. Phys. Condens. Matter* **19,** 183201 (2007).
37. Tserkovnyak, Y., Brataas, A. & Bauer, G. E. W. Spin pumping and magnetization dynamics in metallic multilayers. *Phys. Rev. B* **66,** 224403 (2002).
38. Tserkovnyak, Y., Brataas, A., Bauer, G. E. W. & Halperin, B. I. Nonlocal magnetization dynamics in ferromagnetic heterostructures. *Rev. Mod. Phys.* **77,** 1375–1421 (2005).
39. Mosendz, O. *et al.* Detection and quantification of inverse spin Hall effect from spin pumping in permalloy/normal metal bilayers. *Phys. Rev. B* **82,** 214403 (2010).
40. Garate, I. & Franz, M. Inverse spin-galvanic effect in the interface between a topological insulator and a ferromagnet. *Phys. Rev. Lett.* **104,** 146802 (2010).
41. Yokoyama, T., Zang, J. & Nagaosa, N. Theoretical study of the dynamics of magnetization on the topological surface. *Phys. Rev. B* **81,** 241410(R) (2011).
42. Yokoyama, T. Current-induced magnetization reversal on the surface of a topological insulator. *Phys. Rev. B* **84,** 113407 (2011).
43. Tserkovnyak, Y. & Loss, D. Thin-film magnetization dynamics on the surface of a topological insulator. *Phys. Rev. Lett.* **108,** 187201 (2012).
44. Mahfouzi, F., Nagaosa, N. & Nikolic, B. K. Spin-orbit coupling induced spin-transfer torque and current polarization in topological-insulator/ferromagnet vertical heterostructures. *Phys. Rev. Lett.* **109,** 166602 (2012).
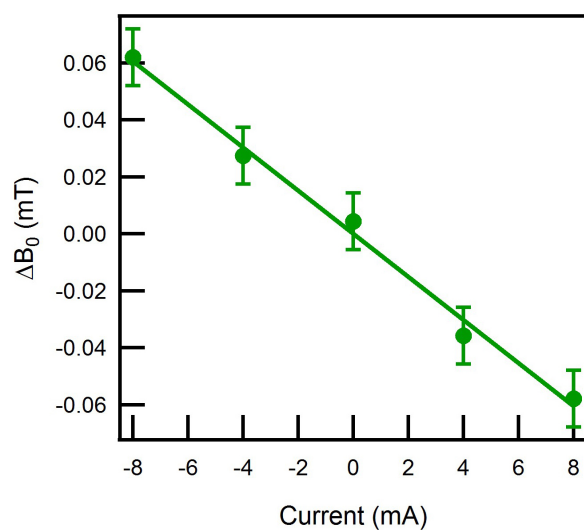
Extended Data Figure 1 | Change in resonant field as a function of direct current. The data correspond to an 8 nm Bi_2Se_3/16 nm permalloy device with dimensions 50 μm × 15 μm, averaged over frequencies between 6 and 10 GHz. Error bars, 1 s.d.
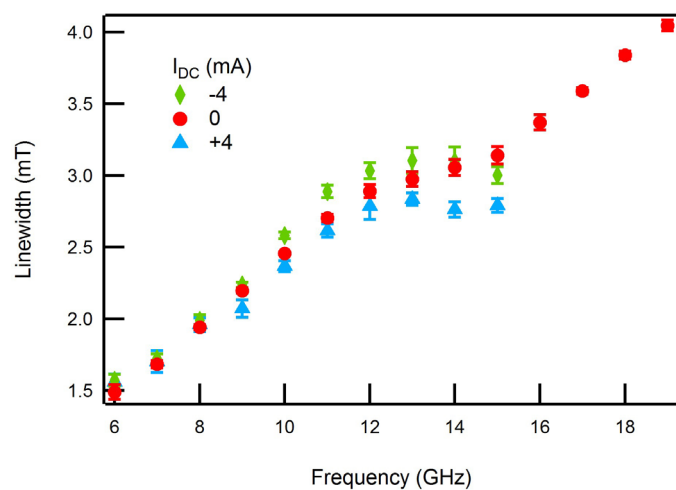
Extended Data Figure 2 | Measured linewidth verses frequency. The data correspond to an 8 nm Bi_2Se_3/16 nm permalloy device with dimensions 50 µm × 10 µm at different direct currents. Error bars, 1 s.d.
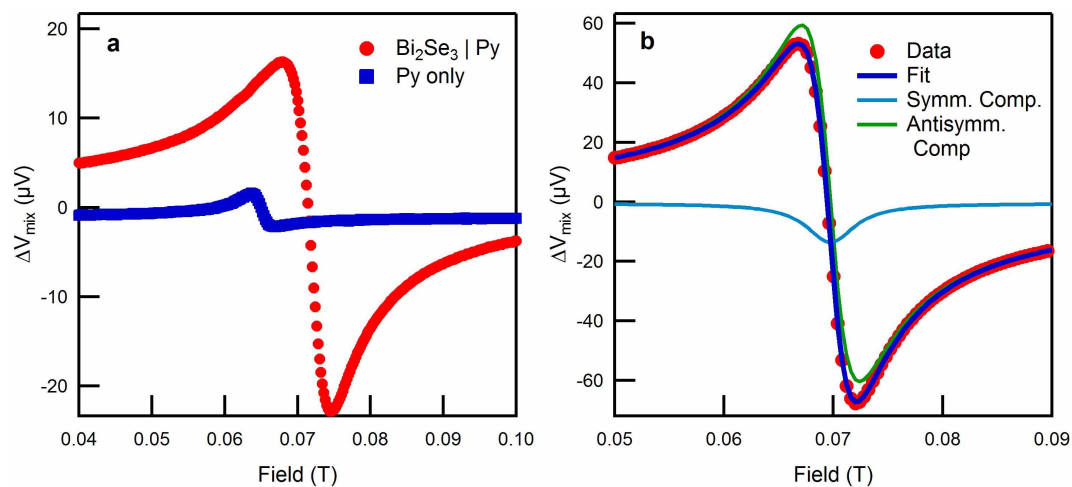
Extended Data Figure 3 | Control experiments. a, A comparison between the ST-FMR signals measured for two 50 μm × 15 μm devices at 8 GHz, one with 8 nm Bi_2Se_3/16 nm permalloy and the other a single layer of 16 nm permalloy. The absorbed radio-frequency power was 5 dBm and $\varphi = 45°$ in both cases. **b**, ST-FMR measurement for a 6 nm Pt/16 nm permalloy device with dimensions 80 μm × 24 μm. The absorbed power was 5 dBm and $\varphi = 45°$.
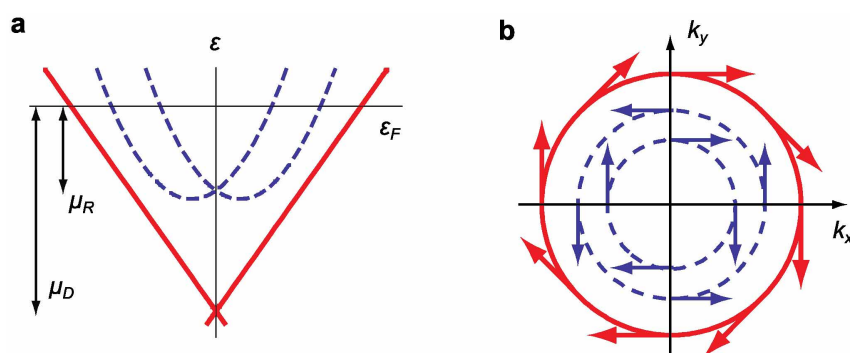
Extended Data Figure 4 | Schematic band structures. a, A schematic band structure with a Dirac surface state and a Rashba-split 2DEG as observed in refs 26, 32. **b**, The corresponding spin angular momentum structure at the Fermi energy.
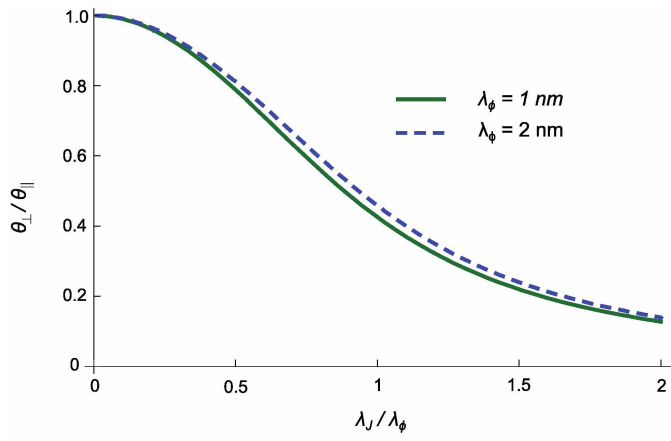
Extended Data Figure 5 | The predicted value of the out-of-plane spin torque ratio relative to the in-plane spin torque ratio. $\theta_\perp/\theta_\parallel$ as a function of λ_J/λ_ϕ, for $\lambda_{sf} = 5$ nm.
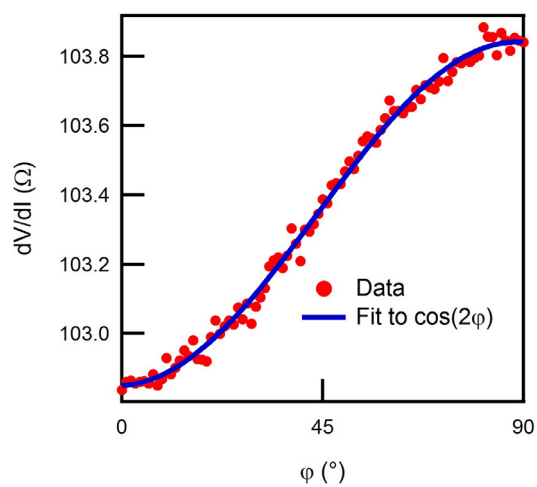
Extended Data Figure 6 | Anisotropic magnetoresistance calibration. The data correspond to an 8 nm Bi_2Se_3/16 nm permalloy device with dimensions $50\,\mu m \times 15\,\mu m$, and were measured at room temperature.

LETTER

A shift of thermokarst lakes from carbon sources to sinks during the Holocene epoch

K. M. Walter Anthony[1], S. A. Zimov[2], G. Grosse[3]†, M. C. Jones[1,4], P. M. Anthony[1], F. S. Chapin III[5], J. C. Finlay[6], M. C. Mack[7], S. Davydov[2], P. Frenzel[8] & S. Frolking[9]

Thermokarst lakes formed across vast regions of Siberia and Alaska during the last deglaciation and are thought to be a net source of atmospheric methane and carbon dioxide during the Holocene epoch[1–4]. However, the same thermokarst lakes can also sequester carbon[5], and it remains uncertain whether carbon uptake by thermokarst lakes can offset their greenhouse gas emissions. Here we use field observations of Siberian permafrost exposures, radiocarbon dating and spatial analyses to quantify Holocene carbon stocks and fluxes in lake sediments overlying thawed Pleistocene-aged permafrost. We find that carbon accumulation in deep thermokarst-lake sediments since the last deglaciation is about 1.6 times larger than the mass of Pleistocene-aged permafrost carbon released as greenhouse gases when the lakes first formed. Although methane and carbon dioxide emissions following thaw lead to immediate radiative warming, carbon uptake in peat-rich sediments occurs over millennial timescales. We assess thermokarst-lake carbon feedbacks to climate with an atmospheric perturbation model and find that thermokarst basins switched from a net radiative warming to a net cooling climate effect about 5,000 years ago. High rates of Holocene carbon accumulation in 20 lake sediments (47 ± 10 grams of carbon per square metre per year; mean $\pm$ standard error) were driven by thermokarst erosion and deposition of terrestrial organic matter, by nutrient release from thawing permafrost that stimulated lake productivity and by slow decomposition in cold, anoxic lake bottoms. When lakes eventually drained, permafrost formation rapidly sequestered sediment carbon. Our estimate of about 160 petagrams of Holocene organic carbon in deep lake basins of Siberia and Alaska increases the circumpolar peat carbon pool estimate for permafrost regions by over 50 per cent (ref. 6). The carbon in perennially frozen drained lake sediments may become vulnerable to mineralization as permafrost disappears[7–9], potentially negating the climate stabilization provided by thermokarst lakes during the late Holocene.

Approximately 30% of global permafrost carbon is concentrated in 7% of the permafrost region (about 1.32 million km^2) in North Siberia, Alaska, and northwest Canada[6] (Extended Data Fig. 1). Here icy, organic-rich, silt-dominated sediments, termed yedoma, were deposited to an average depth of 25 m in unglaciated regions during the late Pleistocene glacial period[10]. The average organic carbon contents of frozen yedoma, typically 2%–5% (refs 5 and 11), are higher and more decomposable than in most thawed mineral soils because fresh organic inputs from the tundra–steppe ecosystem were buried and frozen into rapidly accumulating sediment[10].

During yedoma sedimentation, massive ice wedges grew as seasonal melt water migrated into frost cracks and refroze. When these ice wedges melt under the warmer Holocene climate, the ground subsides (thermokarst), forming deep lakes (10–30 m). Microbes decompose thawed yedoma organic matter in anaerobic lake bottoms, releasing methane, a potent greenhouse gas, which further accelerates permafrost thaw in a positive feedback cycle[1]. These deep, methane-emitting lakes expand until thaw creates a drainage channel. Partial lake drainage lowers lake water level, slowing thermokarst and stimulating growth of benthic mosses and other plants. When lakes drain completely, lakebed sediments, rich with the remains of moss and other organic matter, freeze and reform permafrost (Fig. 1).

Since the last deglaciation (the past 14.7 thousand years, kyr), about 70% of all yedoma area thawed beneath thermokarst lakes and streams (Methods). While thermokarst released a fraction of yedoma carbon to the atmosphere as greenhouse gases[2–4] and a lesser fraction to downstream export (Supplementary Information section 1.5), atmospheric carbon dioxide was absorbed by thermokarst basins through assimilation by contemporary plants. Plant remains are sequestered in newly formed Holocene*-aged permafrost of deep, drained thermokarst-lake basins (alases). Holocene* denotes the Holocene (11.7 kyr ago to present) and pre-Holocene deglacial (14–11.7 kyr ago) period of alas formation (Supplementary Information section 1.1). The extent to which carbon uptake by thermokarst lakes and alases offsets their greenhouse gas emissions has not been determined.

In this study we used field observations of Siberian permafrost exposures, radiocarbon dating, spatial analyses and modelling to (1) quantify the thermokarst-basin carbon stock and (2) document the role of thermokarst-lake carbon accumulation in compensating for greenhouse gas emissions from permafrost thaw in the Holocene.

At 49 widely dispersed steep-bluff sites (up to 25 m high) along rivers and coasts in North Siberian tundra and boreal forests, we quantified permafrost carbon pools in refrozen sediments of alases (Methods, Fig. 2). We compared alas carbon stocks, ^{14}C ages, and macrofossil composition to those of the adjacent undisturbed yedoma (up to 46 m high). We found that yedoma that thawed beneath lakes during the Holocene* and subsequently refroze when the lakes drained lost 28% $\pm$ 12% (mean $\pm$ standard error (s.e.), Extended Data Table 2) of its organic carbon to greenhouse gas production.

Basal dates compiled from alases in North Siberia, Alaska and northwest Canada (collectively called Beringia) indicate widespread thermokarst-lake formation around 14–9 kyr ago (Fig. 3a, Supplementary Table 2), generating a major northern source of atmospheric methane during deglaciation[2,4]. During subsequent millennia, mosses and other plants growing in and around lakes assimilated atmospheric carbon and were buried and preserved in lake sediments, mitigating the loss of permafrost-derived carbon from thermokarst lakes. Our conservative estimate for the mass of Holocene* organic carbon accumulation in Beringian alases (159 ± 24 Pg over 0.93 million km^2) is 1.6 times larger than the mass of Pleistocene-aged yedoma carbon that was released by methane and carbon dioxide production in thawed sediments beneath the same lakes (100 ± 34 Pg) (Fig. 3d; Extended Data Table 3). Additional methane generated from decomposition of Holocene* organic material was also emitted (Supplementary Information section 1.6.3).

[1]Water and Environmental Research Center, University of Alaska, Fairbanks, Alaska 99775-5860, USA. [2]Northeast Scientific Station, Pacific Institute for Geography, Far-East Branch, Russian Academy of Sciences, Cherskii 678830, Russia. [3]Geophysical Institute, University of Alaska, Fairbanks, Alaska 99775-7320, USA. [4]US Geological Survey, Reston, Virginia 20192, USA. [5]Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska 99775-7000, USA. [6]Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, Minnesota 55108, USA. [7]Department of Biology, University of Florida, Gainesville, Florida 32611, USA. [8]Max Planck Institute for Terrestrial Microbiology, Marburg 35043, Germany. [9]Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire 03824-3525, USA. †Present address: Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam 14473, Germany.
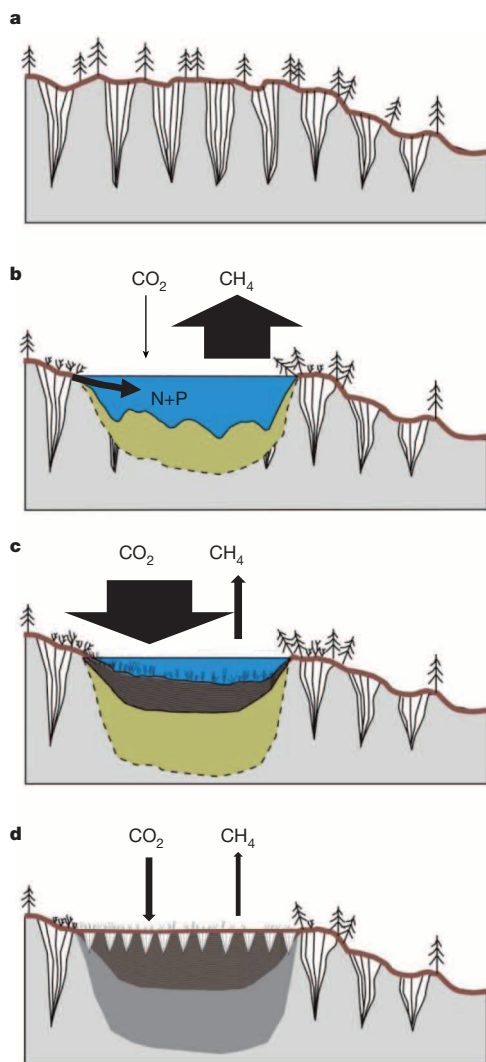
Figure 1 | Carbon cycling during the development of deep thermokarst lakes. a, Yedoma with massive Pleistocene ice wedges. **b,** Thermokarst-lake expansion (thaw bulb shown in yellow; thaw boundary shown as a dotted line) accompanied by Pleistocene-aged CH_4 emissions and release of N and P from yedoma into lakes, stimulating aquatic productivity and CO_2 uptake, which offset Pleistocene-aged CO_2 emissions from yedoma decay. **c,** Partially drained lake with atmospheric CO_2 uptake by plants forming thick Holocene* organic carbon deposits (brown) exceeding CH_4 emissions from contemporary organic matter decay. **d,** Refreezing of remaining Pleistocene and Holocene* carbon in sediments following complete lake drainage (new ice wedges are shown as triangles), with peatland-type CH_4 emissions and CO_2 uptake. The thicknesses of CH_4 and CO_2 arrows are scaled by relative magnitude on a carbon-mass basis (Supplementary Information section 1.6).

Since greenhouse gas emissions and carbon sequestration have counteractive effects on climate (warming versus cooling, respectively), the radiative impacts of both processes must be scaled for comparison. We developed simple deglacial-to-present thermokarst-basin carbon flux trajectories (Fig. 3b) based on estimates of contemporary methane flux (Supplementary Information section 1.6.3), total yedoma carbon lost as carbon dioxide and methane (100 ± 34 Pg), total accumulated Holocene* carbon (159 ± 24 Pg), and thermokarst-lake initiation dates. Flux trajectories were used as input to a simple atmospheric perturbation model developed originally to assess northern peatland carbon feedbacks to climate in the Holocene[12] (Extended Data Table 4). Thermokarst lakes caused a net climate warming of 0.06 ± 0.03 W m^{-2} at the peak of their formation during deglaciation (Fig. 3c), driven primarily by methane release from thawed, decaying yedoma. The peak climate impact of yedoma

lakes was much larger than that of all northern peatlands in the early Holocene[12]. Owing to high carbon accumulation in existing basins and a slowdown of lake formation, thermokarst-lake impact on climate switched from net warming to net cooling about 5 kyr ago. Since that time these carbon-sequestering lake basins have caused an increasing net cooling to the present-day net radiative forcing of -0.06 ± 0.02 W m^{-2} (Fig. 3c).

Following drainage, the peat-rich sediments of these lake basins refroze, sequestering a large, Holocene* C pool in permafrost. This deep-lake carbon, now part of the terrestrial carbon stock, was not included in previous global carbon inventories because alas deposits had not been systematically sampled for this purpose[6,13]. Rather, past approaches assumed homogeneous yedoma carbon concentrations for deep frozen deposits in the region (Supplementary Information section 3.5). This newly recognized alas carbon pool (159 ± 24 Pg) is more than twice as large as the West Siberian Lowland peat-carbon stock (70 Pg; ref. 14) and increases the existing estimate of northern circumpolar permafrost-zone peatland carbon (277 Pg; ref. 6) by 56%.

Past syntheses have suggested that climatic controls limit carbon accumulation in lakes north of 62° N to less than 10 g m^{-2} yr^{-1} during the Holocene[15]. Radiocarbon dating of terrestrial plant macrofossils buried in our Siberian study-lake sediments revealed exceptional rates of long-term carbon accumulation (47 ± 10 g C m^{-2} yr^{-1}, mean $\pm$ s.e., $n = 20$ lakes). These rates are about 2.5-fold and 5-fold higher than the global averages for northern peatlands (18.6 g C m^{-2} yr^{-1}; ref. 16) and lakes (4.5–14 g C m^{-2} yr^{-1}; ref. 17) (Fig. 4), respectively. We found similarly high carbon accumulation (30–182 g C m^{-2} yr^{-1}) when we reconstructed published ^{14}C-dated alas profiles in other yedoma subregions; similar rates were also found in deep, arctic Alaska lakes (Supplementary Table 3).

We attribute the high carbon accumulations in our Siberian study lakes to: (1) thermokarst-related shore erosion and deposition of terrestrial organic matter in lake bottoms; (2) high aquatic productivity enhanced by nutrient supply from thawing yedoma; and (3) the unique preservation conditions in deep thermokarst lakes. In addition to the initial surface layer of organic soils slumping into lakes as topographic changes began, the productivity of terrestrial communities surrounding lakes increased when thermokarst lakes formed. We observed enhanced shrub growth along modern lake margins, probably resulting from the protected environment created by thaw subsidence (for example, enhanced snow accumulation and wind protection); deepening of the active layer, which mobilizes nutrients from thawed soils; and the concave shape of thermokarst features, which traps nutrient and water flow from the surrounding watershed[18]. As the lake margins expand, shrubs fall into the lakes.

Macrofossil analysis and ^{14}C dating of modern and palaeolake sediments indicated that upland plants, littoral sedges and benthic brown mosses contributed nearly equally to the macrofossil remains of the deep lake deposits (Fig. 2a); however, we observed the fastest rates of carbon accumulation (89 ± 15 g C m^{-2} yr^{-1}, mean $\pm$ s.e., $n = 8$, maximum 185 g C m^{-2} yr^{-1}) in sequences dominated by thick, *in situ,* benthic-moss peat (Supplementary Information section 3.2). Given their capacity to flourish under low temperature and light[19], brown mosses (Amblystegiaceae) are the dominant macrophyte in arctic lakes[20], including the North Siberian thermokarst lakes. We observed such mosses in all present-day yedoma lakes, but they were most prolific along floating mats in partially drained, yet still deep (8–12 m) boreal yedoma lakes. Aquatic mosses are particularly responsive to phosphorus fertilization[20,21] and nutrient release from permafrost thaw on lakeshores[22]. In our study region, thawing yedoma releases nitrogen and phosphorus (Supplementary Information section 1.8, Extended Data Table 5), leading to high hypolimnetic orthophosphate concentrations (mean 233 µg P per litre), and moss-tissue nitrogen ($2.0 \pm 0.22\%$ dry weight) and phosphorus ($0.12 \pm 0.01\%$ dry weight) concentrations above critical levels for moss growth[20]. Nutrient concentrations in North Siberian yedoma lakes were one to three orders of magnitude higher than in shallow, non-yedoma thermokarst lakes
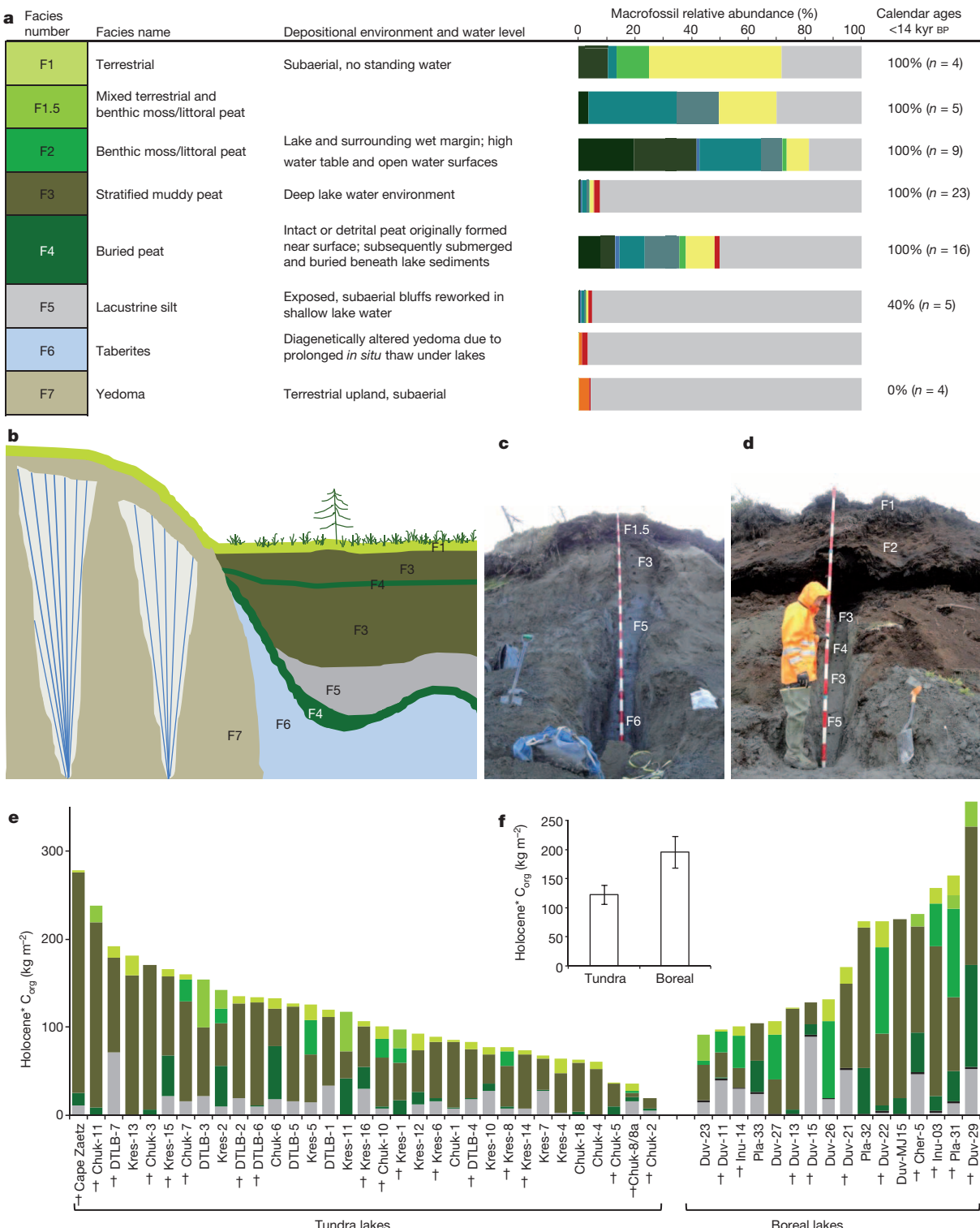
Figure 2 | Facies description and carbon contents in the deep thermokarst-lake landscape. a, Facies depositional environment; relative abundance of macrofossils (dark green, aquatic moss; mid-green, wet moss; light blue, other aquatic; turquoise, wet sedge; blue, other emergent mosses; light green, unidentified; yellow, terrestrial; orange, dry graminoid; red, detritus in silt; grey, inorganic) (Supplementary Table 1) scaled by the mean organic matter content of the facies; and fraction of macrofossils with ^{14}C ages <14 kyr ago in permafrost exposures (Extended Data Table 1). Facies were determined in the

field based on physical properties; subsequent ^{14}C dating confirmed facies ages (Pleistocene versus Holocene*). The schematic (**b**) and photographs (**c**, **d**) show examples of facies in cross-section; pole marks are 20 cm. Coloured bars in **e** indicate Holocene* organic carbon stocks by facies observed in 49 refrozen, thermokarst exposures (see **a**; black, wood; dagger, fully exposed). **f**, Boreal-zone thermokarst basins accumulated more Holocene* organic carbon (mean ± s.e.; 196 ± 27 kg C m^{-2}, n = 10) than tundra basins (122 ± 16 kg C m^{-2}, n = 18; one-sided t-test, P < 0.05).

in the same study region (Extended Data Fig. 3) and deep, non-yedoma lakes in Alaska (Extended Data Fig. 4), where mosses are less abundant[23].

Cold and anoxic conditions fostering organic matter preservation[7,8] also lead to high sediment carbon accumulation. Strong stratification resulting from high solute concentrations[24] (indicated by high specific
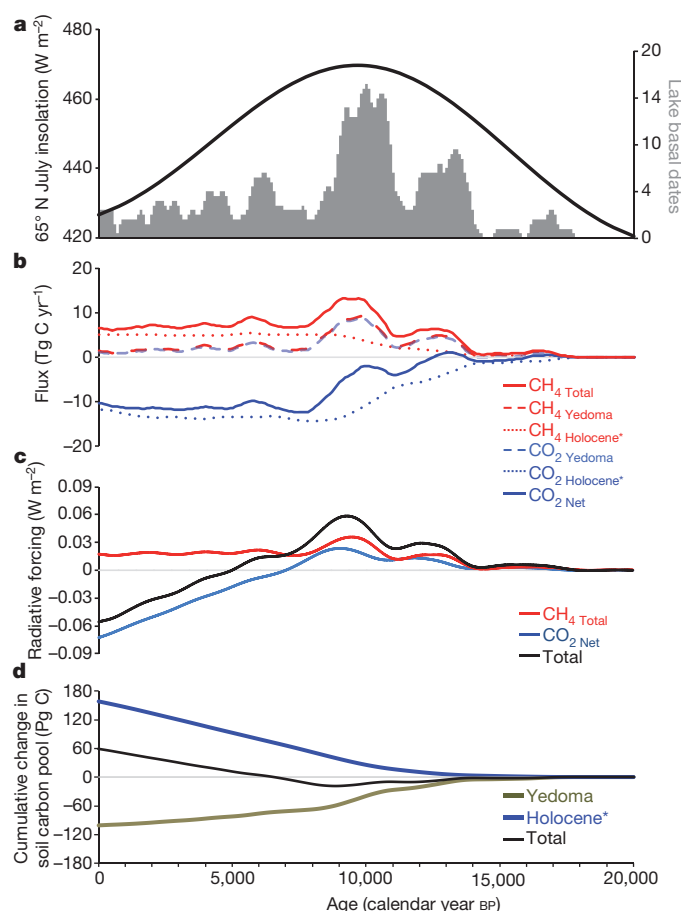
Figure 3 | Thermokarst-lake carbon cycling dynamic since the last deglaciation. a, The increase in arctic insolation[30] coincided with widespread thermokarst-lake formation during deglaciation (Supplementary Information section 1.6.1, Supplementary Table 2). **b**, Carbon flux trajectories for yedoma-region thermokarst basins determined by basal date frequency in **a** (CO_2-C (the carbon component of carbon dioxide) uptake by peat formation, negative flux; CO_2-C emission from thawed yedoma decay and CH_4-C (the carbon component of methane) emissions, positive flux). The solid red line in **b** is the sum of CH_4-C emissions from decay of thawed yedoma deposits (dashes) and younger organic matter termed Holocene* (Supplementary Information section 1.1) in thermokarst basins (dots). Identical yedoma-derived CH_4-C and CO_2-C emission curves are based on methanogenesis stoichiometry (Supplementary Information section 1.6.2). **c**, Radiative forcing due to atmospheric perturbations in CH_4 and CO_2 concentration for flux trajectories shown in **b**. **d**, Cumulative changes in the yedoma-region permafrost carbon pool due to loss of yedoma carbon to the atmosphere by thermokarst-lake formation (negative) and atmospheric carbon uptake and burial by the same lakes (positive). See Supplementary Information section 1.6 and Extended Data Table 3 for detailed methods and uncertainties.

conductivity and dissolved organic carbon; Extended Data Figs 3 and 4), protection from wind mixing by steep banks and small surface-area to volume ratios, and proximity to permafrost maintained unusually cold lake bottom temperatures (down to 1.5 °C year round). When lakes drain completely, sediment organic matter rapidly freezes (on a time-scale of decades)[25], stabilizing this carbon in permafrost.

Permafrost temperatures are increasing in Siberia, as in most of the Arctic[26]. Global annual air temperature increases as little as 1.5 °C, less than the projected 1.9 °C–3.7 °C air temperature increase by 2100 AD[27], should lead to pronounced reductions in permafrost stability[28]. Our results suggest that if new, deep thermokarst lakes continue to form by localized permafrost thaw, then primary productivity and sedimentation in lakes, particularly tundra lakes where aquatic productivity is expected
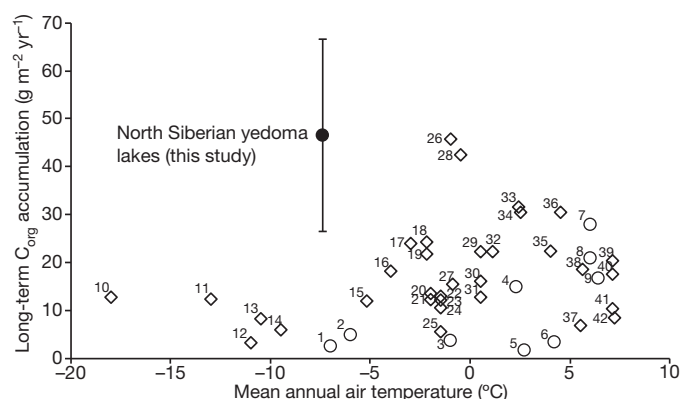

Figure 4 | Comparison of long-term organic carbon accumulation rates among northern lakes and peatlands by mean annual temperature. Despite cold temperatures in the North Siberia yedoma region, thermokarst lakes there (closed circle, mean and 95% confidence interval) accumulated organic carbon faster than other northern lakes (open circles labelled 1–9) and peatlands (open diamonds labelled 10–42; except the West Siberian Lowlands, labelled 26 and 28), 228 European lakes (5.6 g m^{-2} yr^{-1}; ref. 15) and global lakes (4.5–14 g m^{-2} yr^{-1}; ref. 17). See Supplementary Information section 3.1 for regional data references.

to increase (Supplementary Information section 3.3), will compensate for greenhouse emissions over millennial timescales if a permafrost-forming climate persists in the region.

Thermokarst lakes depend on permafrost for an impermeable base that constrains water drainage[29]. Widespread permafrost loss, predicted as early as 2100 in some yedoma subregions[9,27], will ultimately result in reduced lake and wetland abundance caused by drainage and drying[8], facilitating rapid sediment carbon decomposition[7]. Even partial mineralization of the alas carbon pool, together with the surrounding Pleistocene-yedoma carbon that remains (all together about 450 Pg C; Extended Data Fig. 5 and Extended Data Table 3) may reverse the important role that thermokarst lakes have had in stabilizing climate for the past 5,000 years.

METHODS SUMMARY

In the field we described 49 drained, refrozen thermokarst-basin (alas) profiles and six yedoma profiles exposed along rivers and coasts in the North Siberian yedoma region (Extended Data Fig. 1). At each site we collected samples and measured the vertical distribution and cumulative thickness of individual stratigraphic facies F1–F7 (Fig. 2a and Extended Data Fig. 2, Supplementary Information section 1.3). Exposure thickness averaged 5.8 m in alases (range 1.25–20 m; Extended Data Table 1) and 31.5 m (range 14–40 m) in yedoma. From stratigraphy and our under-standing of geomorphologic and palaeoenvironmental processes associated with the facies, we calculated the Pleistocene and Holocene* carbon accumulated in each profile using three independent approaches: (1) macrofossils (Fig. 2 and Supplementary Table 1), (2) radiocarbon dates (Fig. 2), and (3) carbon mass-balance equations (Supplementary Information section 1.4).

We estimated the yedoma carbon stock (325 ± 69 kg C m^{-2}) as the product of the mean yedoma organic carbon bulk density (26.0 ± 1.5 kg C m^{-3}) measured in eight subregions between the Kolyma Lowlands and Laptev Sea by us in this study (Extended Data Table 2) and ref. 5, an ice wedge volume of 50% ± 3% (refs 10 and 11), and an average deposit thickness of 25 ± 5 m based on broader regional estimates (Extended Data Tables 1 and 3). We calculated Holocene* carbon stocks in 49 individual refrozen thermokarst exposures as the sum of products of organic carbon bulk density and thickness measured for facies F1–F5 minus Holocene ice-wedge volume (10% of the 2-m surface layer) (Fig. 2e).

The deep, thermokarst-lake carbon pool estimate (159 ± 24 Pg) is the tundra/ boreal area-weighted average Holocene* carbon stock measured in 28 fully exposed, refrozen thermokarst study sites (172 ± 19 kg C m^{-2}; Fig. 2f) extrapolated to the spatial extent of deep thermokarst-lake basins and thermoerosional gullies in the region (925,400 ± 93,000 km^2) (Extended Data Table 3). Assumptions, uncertainty analysis, and detailed methodology are described in the online-only Methods and the Supplementary Methods.

1. Walter, K. M., Zimov, S. A., Chanton, J. P., Verbyla, D. & Chapin, F. S., III. Methane bubbling from Siberian thaw lakes as a positive feedback to climate warming. *Nature* **443,** 71–75 (2006).
2. Walter, K. M., Edwards, M. E., Grosse, G., Zimov, S. A. & Chapin, F. S. III. Thermokarst lakes as a source of atmospheric CH_4 during the last deglaciation. *Science* **318,** 633–636 (2007).
3. Petrenko, V. V. et al. $^{14}CH_4$ measurements in Greenland ice: investigating last glacial termination CH_4 sources. *Science* **324,** 506–508 (2009).
4. Brosius, L. S. et al. Using the deuterium isotope composition of permafrost meltwater to constrain thermokarst lake contributions to atmospheric CH_4 during the last deglaciation. *J. Geophys. Res.* **117,** G01022 (2012).
5. Schirrmeister, L. et al. Fossil organic matter characteristics in permafrost deposits of the northeast Siberian Arctic. *J. Geophys. Res.* **116,** G00M02 (2011).
6. Tarnocai, C. et al. Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23,** GB2023 (2009).
7. Maltby, E. & Immirzi, P. Carbon dynamics in peatlands and other wetland soils: regional and global perspectives. *Chemosphere* **27,** 999–1023 (1993).
8. Avis, C. A., Weaver, A. J. & Meissner, K. J. Reduction in areal extent of high-latitude wetlands in response to permafrost thaw. *Nature Geosci.* **4,** 444–448 (2011).
9. Slater, A. G. & Lawrence, D. M. Diagnosing present and future permafrost from climate models. *J. Clim.* **26,** 5608–5623 (2013).
10. Zimov, S. A., Schuur, E. A. G. & Chapin, F. S. Permafrost and the global carbon budget. *Science* **312,** 1612–1613 (2006).
11. Strauss, J. et al. The deep permafrost carbon pool of the Yedoma region in Siberia and Alaska. *Geophys. Res. Lett.* **40,** 6165–6170 (2013).
12. Frolking, S. & Roulet, N. T. Holocene radiative forcing impact of northern peatland carbon accumulation and methane emissions. *Glob. Change Biol.* **13,** 1079–1088 (2007).
13. Hugelius, G. et al. The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth Syst. Sci. Data* **5,** 3–13 (2013).
14. Smith, L. C. et al. Siberian peatlands a net carbon sink and global methane source since the early Holocene. *Science* **303,** 353–356 (2004).
15. Kastowski, M., Hinderer, M. & Vecsei, A. Long-term carbon burial in European lakes: Analysis and estimate. *Glob. Biogeochem. Cycles* **25,** GB3019 (2011).
16. Yu, Z., Beilman, D. W. & Jones, M. C. in *Carbon Cycling in Northern Peatlands* (eds Baird, A. J., Belyea, L. R., Comas, X., Reeve, A. S. & Slater, L. D.) Geophysical Monograph Series 184 (AGU, 2009).
17. Tranvik, L. J. et al. Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol. Oceanogr.* **54,** 2298–2314 (2009).
18. Lantz, T. C., Kokelj, S. V., Gergel, S. E. & Henry, G. H. R. Relative impacts of disturbance and temperature: persistent long-term changes in microenvironment and vegetation in retrogressive thaw slumps. *Glob. Change Biol.* **15,** 1664–1675 (2009).
19. Welch, H. E. & Kalff, J. Benthic photosynthesis and respiration in Char Lake. *J. Fish. Res. Board Can.* **31,** 609–620 (1974).
20. Riis, T., Olesen, B., Katborg, C. K. & Christoffersen, K. S. Growth rate of an aquatic bryophyte (*Warnstorfia fluitans* (Hedw.) Loeske) from a high arctic lake: effect of nutrient concentration. *Arctic* **63,** 100–106 (2010).
21. Bowden, W. B., Finlay, J. C. & Maloney, P. E. Long-term effects of PO_4 fertilization on the distribution of bryophytes in an arctic stream. *Freshwat. Biol.* **32,** 445–454 (1994).
22. Mesquita, P. S., Wrona, F. J. & Prowse, T. D. Effects of retrogressive permafrost thaw slumping on sediment chemistry and submerged macrophytes in Arctic tundra lakes. *Freshwat. Biol.* **55,** 2347–2358 (2010).
23. Hershey, A. Effects of predatory sculpin on the chironomid communities in an arctic lake. *Ecology* **66,** 1131–1138 (1985).
24. Houser, N. J. Water color affects the stratification, surface temperature, heat content, and mean epilimnetic irradiance of small lakes. *Can. J. Fish. Aquat. Sci.* **63,** 2447–2455 (2006).
25. Jones, B. M. et al. Modern thermokarst lake dynamics in the continuous permafrost zone, northern Seward Peninsula, Alaska. *J. Geophys. Res.* **116,** G00M03 (2011).
26. Romanovsky, V. E. et al. Thermal state of permafrost in Russia. *Permafrost Periglacial Process.* **21,** 136–155 (2010).
27. Stocker, T. F. et al. Technical summary. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) (Cambridge Univ. Press, 2014).
28. Vaks, A. et al. Speleothems reveal 500,000-year history of Siberian permafrost. *Science* **340,** 183–186 (2013).
29. Smith, L. C., Sheng, Y. & MacDonald, G. M. A first pan-Arctic assessment of the influence of glaciation, permafrost, topography and peatlands on northern hemisphere lake distribution. *Permafrost Periglacial Process.* **18,** 201–208 (2007).
30. Berger, A. & Loutre, M. F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10,** 297–317 (1991).

Author Contributions K.M.W.A. had primary responsibility for study design, field work, laboratory measurements, data analysis, interpretation and writing. S.A.Z. co-designed the study and contributed substantially to data interpretation. M.C.J., G.G., P.M.A. and F.S.C. contributed to project planning, field and laboratory work, and interpretation of results. M.C.J. provided expertise in macrofossil identification. G.G. conducted spatial analyses. K.M.W.A., M.C.M., J.C.F. and S.D. conducted laboratory analyses of lake water samples and ice wedges, and designed and implemented the component of terrestrial vegetation and soil nutrient cycling. P.F. conducted anaerobic laboratory incubations. S.F. created the atmospheric model for radiative forcing calculations. All authors contributed to the revision and integration of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M.W.A. (kmwalteranthony@alaska.edu).

METHODS

Permafrost alas and yedoma exposures. Thermokarst profiles represented refrozen lake-sediment sequences and overlying drained-lake basin soils (alases) at all sites, except Duv-26, which was a thermoerosional gully associated with a former stream. Fieldwork was conducted in August (of 2009 and 2011), when seasonal thaw is greatest. At each study outcrop we removed seasonally thawed sediment with shovels to expose permafrost in cross-section along a vertical profile (Fig. 2). We studied exposures with the steepest bluff angles to avoid locations where permafrost may have thawed, slumped and re-frozen since formation. Bluff angle was a function of present-day erosion dynamics along rivers and coasts into the palaeolake sediment sequences; it was not a function of location in lake basins. Detailed mapping of lake bed morphology in modern yedoma thermokarst lakes[31] and palaeolakes (this study) has revealed that these lake beds are not flat, with the exception of old or large lakes where erosion and sedimentation have partially worn down baydjarakhs (thermokarst mounds) and completely filled in thermokarst depressions. Relatively high relief in actively expanding yedoma thermokarst occurs at the basin scale, but also at the finer scale of the former ice-wedge network (approximately 10–12 m)[32–34]. At the basin scale, yedoma thermokarst lakes have steep bluffs (often >10 m), particularly along expanding thermokarst margins where ground ice rapidly melts, leading to subsidence and high relief. At a finer scale, the bottom of an actively expanding thermokarst lake has a morphology similar to an egg carton, with micro-relief among baydjarakhs often exceeding 1 m. In contrast to core drilling and sampling, studying permafrost exposures has the advantage of providing a broader perspective on the distribution of sedimentary facies because it is possible to visually trace horizons identified in exposures across lake basins. In palaeolakes, we often saw the undulating pattern of baydjarakhs at the base of the sedimentary sequence, which later formed a flatter surface across basins as lakes matured and filled in with peat-rich sediments (Fig. 1).

The use of a single sampling location in lake basins with strong sediment focusing (accumulation of fine particles in deep parts of a lake basin) has the potential to lead to false conclusions about climate-driven sediment deposition dynamics and inflated carbon accumulation rates if sedimentation rates are improperly scaled. Multiple, distributed cores, geophysical methods for cross-basin scaling of sedimentation, and focusing-corrected carbon burial rates have been used to avoid such false conclusions[35–37]. Since the bluffs we studied were located at various, non-uniform positions within basins (that is, not always in the lake centre or at sites with a particular deposit thickness), there was no bias in our sampling towards environments of sediment focusing. This also implies that among the sets of boreal and tundra exposures we studied, various depositional environments in the former lakes (that is, near-shore, transitional, and pelagic zones) are represented.

With shovels we exposed cross-sections of Holocene and Pleistocene deposits from the ground surface down to depths ranging from 1.25 m to 20 m below the surface in the 49 alas sites (exposure depth: mean 5.6 m, median 4.8 m, range 1.25–20 m; Extended Data Table 1). Among the 28 exposures where we had high confidence that we had dug close to the base of the Holocene* deposits, the mean thickness of sediments containing Holocene* carbon was 5.8 m (median 5.0 m). Average alas bluff height, including uncovered deposits beneath our study exposures, was 13 ± 1 m (s.e.), $n = 49$. Our map-based analysis of undisturbed yedoma deposit thickness surrounding the alas exposures yielded 38 ± 2 m (s.e.), $n = 17$.

At each alas exposure site we measured the vertical distribution and cumulative thickness of each of facies F1–F6 (Supplementary Information section 1.3). In a subset of nine alases we measured the thickness of surface peat deposits along cross-basin transects with 5 to 15 measurement sites per basin. We found that the basin mean peat thicknesses were not statistically different from the point measurements at single exposures (Wilcoxon signed rank test, $P = 0.6$). This demonstrates that there was no bias in our field sampling locations.

In addition to alases, we sampled yedoma (F7) profiles representing the Pleistocene permafrost of non-thermokarst surfaces in each of the North Siberia study regions, except Krestovskiy Cape (Extended Data Fig. 1). The sampling depths of yedoma ranged from 1.5 m to 37 m below the ground surface. Yedoma exposure thickness averaged 31.5 m (range 14–40 m). We did not sample the Holocene carbon content of soils overlying the undisturbed Pleistocene-aged yedoma deposits.

After carefully cleaning study exposures with spatulas, we characterized the cryolithology, differentiated the facies, and collected 595 soil samples (up to 39 samples per vertical profile) for laboratory analyses. In 2009, we chipped 212 frozen samples (0.3–0.5 kg) from exposures with hatchets and hammers. In 2011 we used a 75-cm³ hole-saw mounted on a hand-held hammer drill to collect 383 samples of known volume for dry bulk density determination. To our knowledge, this constitutes the largest data set of yedoma-region permafrost samples from which population-level organic carbon bulk density could be calculated on the basis of the products of bulk density and organic matter content measured on the same samples. Alas samples were collected as one individual sample per depth; 2011 yedoma (F7) samples were usually collected in triplicate. Analytical results of triplicate samples were averaged to

avoid pseudo-replication in statistical analyses. We estimated gravimetric ice content of permafrost samples as weight loss after drying at 105 °C to constant weight, expressed as a percentage of fresh weight (wt%). We determined the soil moisture of thawed, active layer samples in alases using the same method. We determined organic and inorganic matter contents of all samples after oven drying by loss-on-ignition at 550 °C and 950 °C for two hours, respectively[38]. On a subset of 63 samples we determined total organic carbon as the difference between total carbon using a Leco CHN analyser and the inorganic carbon component using titration[39]. We estimated organic carbon concentrations in the remainder of the samples based on the relationship between loss-on-ignition (percentage organic matter, P_{om}) and organic carbon concentration (P_{oc}) as follows: $P_{oc} = 0.5544 \times P_{om}$ ($P < 0.0001$; $r^2 = 0.9969$). The organic carbon bulk density (in units of $kg\,C\,m^{-3}$) was calculated as dry bulk density multiplied by percent organic carbon on a per sample basis.

We characterized macrofossil assemblages on 225 field samples collected in alas and yedoma exposures (Supplementary Information section 1.2, Supplementary Table 1). From stratigraphy and our understanding of geomorphologic and palaeoenvironmental processes associated with the facies (Supplementary Information section 1.3), we calculated the Pleistocene and Holocene* carbon accumulated in each profile using three independent approaches: (1) macrofossil assemblages (Fig. 2, Supplementary Table 1), (2) radiocarbon dates (Fig. 2), and (3) carbon mass-balance equations (Supplementary Information section 1.4). We refer to organic matter fixed via photosynthesis in and around the thermokarst basins as 'Holocene*' because the majority (93%) accumulated from 11.7 kyr ago to present (Fig. 3d); however, it should be noted that 7% of the non-yedoma, peat-rich carbon pool indicated by 'Holocene*' actually accumulated in thermokarst basins that formed before the Holocene during the deglacial warming between 14 kyr ago and 11.7 kyr ago. This follows terminology common to peatland studies[12,14]. Without the asterisk, Holocene refers to the period of 11.7 kyr ago to present.

Radiocarbon dating. We obtained accelerator mass spectrometry (AMS) radiocarbon dates on plant macrofossils picked from permafrost exposures and present-day lake sediment cores. Samples were analysed at the W. M. Keck Carbon Cycle AMS Laboratory or at the National Ocean Sciences AMS Facility (Supplementary Tables 2 and 3). All radiocarbon ages were calibrated to calendar ($\pm 2\sigma$) years before present (BP) using Calib 6.0 software (ref. 40).

Yedoma and thermokarst basin areas. We calculated the original Last Glacial Maximum extent of the core yedoma region in land areas exposed today in Beringia (1,322,000 km²) using general maps of yedoma distribution in North and East Siberia (1,141,000 km²) (ref. 41) and a surface geological map of Alaska that shows ice-rich silty deposits (181,000 km²) (ref. 42), which we classify as yedoma. Based on literature data[43–47] and a new assessment of yedoma spatial continuity using high-resolution Siberian geological maps[48], we estimated the fraction of areas with degraded yedoma (about 70%; 925,400 km²) within the original yedoma region land areas, including deep thermokarst-lake basins (50%–60%) and streams and rivers (about 10%–20%), as well as the remaining 30% of undisturbed yedoma that had not degraded since the Pleistocene. Deep thermokarst-lake basins include both present-day lakes as well as drained, refrozen lake basins. Landforms shaped by melting ground ice and fluvial erosion of streams and small rivers are known as 'thermoerosional gullies'. In our calculations we do not include known but smaller yedoma occurrences in valleys of the Yukon Territories (Canada) and Chukotka and Taymyr peninsulas (Russia).

To estimate the current ratio of yedoma located in tundra and boreal forest regions, we used the Circum Arctic Vegetation Map[49] to determine the yedoma-region extent in these two ecoregion categories. We calculated that 32% of the yedoma region is located in today's tundra regions (424,000 km²) and 68% is located in boreal forest regions (898,000 km²). Supplementary Information section 1.7.3 provides an uncertainty analysis of the geospatial analysis.

Pleistocene and Holocene carbon stocks and pool sizes. Parameters for calculating the Pleistocene and Holocene* carbon stocks and pool sizes are shown in Extended Data Tables 2 and 3a. The uncertainty analysis is summarized in Extended Data Table 3b and described in detail in Supplementary Information section 1.7.

We estimated the yedoma carbon stock ($325 \pm 69\,kg\,C\,m^{-2}$) as the product of the mean yedoma organic carbon bulk density ($26.0 \pm 1.5\,kg\,C\,m^{-3}$), measured in eight subregions between the Kolyma Lowlands and Laptev Sea by us in this study (Extended Data Table 2) and ref. 5, an average yedoma facies thickness of 25 ± 5 m, as determined from extensive literature[10,32,50–56], and an ice-wedge volume of $50 \pm 3\%$ (refs 10, 11, 32, 56). The error value of 3% is the approximate standard error of the mean ice wedge volume determined from careful measurements of ten sites[11]: 48% mean, 52% median, $n = 10$, min 35%, max 60%.

We estimated the Pleistocene carbon pool size for the 1.32 million km² yedoma territory of mainland Beringia (excluding continental shelves) at the Last Glacial Maximum (429 ± 101 Pg C) by applying the yedoma carbon stock ($325 \pm 69\,kg\,C\,m^{-2}$) to this area before extensive thermokarst activity occurred and propagating uncertainty associated with mean estimates of carbon bulk density (Extended Data Table

2), yedoma extent (Supplementary Information section 1.7.3), yedoma thickness and ice-wedge volume (Extended Data Table 3). Today the Pleistocene carbon pool size is 34% lower (284 ± 40 Pg C), since 70% of the original Last Glacial Maximum yedoma (area and volume) was degraded by thermokarst processes during the Holocene* and partially eroded by rivers (Supplementary Information section 1.5) or decomposed by microbes living in and beneath water bodies. The present-day Pleistocene carbon pool (284 ± 40 Pg C) is the sum of: (1) the remaining undisturbed yedoma (129 ± 30 Pg C), (2) thawed or previously thawed yedoma in thermokarst-lake basins (141 ± 26 Pg); and (3) thawed or previously thawed yedoma in the beds of thermokarst streams and rivers (14 ± 2 Pg) (Extended Data Table 3).

Upscaling calculations of the Holocene* carbon pool in thermokarst basins is based on the means of individual thermokarst exposure carbon stocks measured in tundra and boreal regions (Fig. 2e). We determined the Holocene* organic carbon stock of each thermokarst study exposure (in units of kg C m^{-2}; Fig. 2e) as the product of the Holocene* organic-carbon bulk density, measured on individual samples, and facies thickness in each exposure minus Holocene ice-wedge volume (10% of the 2-m surface layer). In instances where the thickness of a facies was measured in an exposure, but no representative samples collected, we applied the mean carbon bulk density for that facies calculated from all measured samples (Extended Data Table 1).

The deep, thermokarst-lake carbon pool estimate (159 ± 24 Pg) is the tundra/boreal area-weighted average Holocene* carbon stock measured in 28 fully exposed, refrozen thermokarst study sites (172 ± 19 kg C m^{-2}; Fig. 2f and Extended Data 2) extrapolated to the spatial extent of deep thermokarst-lake basins and thermoerosional gullies in the region ($925,400 \pm 93,000$ km^2). This upscaling assumes that the 28 fully-excavated exposures represent the population of deep thermokarst-lake basin and thermoerosional gully environments in the yedoma region of Beringia (Fig. 2). The carbon stock data for the 28 exposures was normally distributed; a similar carbon accumulation was estimated for alases in other yedoma subregions (Supplementary Table 3). Extended Data Table 3 details our calculations and shows that upscaling based on median carbon stock values (104 kg C m^{-2} tundra, 195 kg C m^{-2} boreal) instead of mean values would result in a 4% difference in the landscape-scale estimate. We included thermoerosional gullies in our spatial extrapolation of Holocene* carbon deposits because we documented that the Holocene carbon accumulation in the thermoerosional gully Duv-26 was similar to that of the deep thermokarst lakes (Fig. 2e). Similar Holocene peat accumulations on top of thawed yedoma deposits have also been observed in thermoerosional gullies in other subregions of yedoma (for example, Laptev Sea)[57–59]. Additional thermokarst lakes formed in extensive, but thinner, yedoma deposits on the flanks of higher-relief topography and on previously exposed continental shelves. These thermokarst sites would increase the Holocene* carbon-stock estimates for thermokarst and their forcing effect on atmospheric carbon dioxide through time, but are not included in the calculations, which are therefore conservative. Our estimate of the thermokarst carbon pool is conservative for additional reasons, which are described in detail in Supplementary Information section 1.7.4.

Radiative forcing impact of deep thermokarst-lake carbon fluxes. We adapted the five-box atmospheric carbon dioxide and one-box atmospheric methane model established for northern peatlands[12] (Extended Data Table 4) to derive the net radiative effect of deep thermokarst basins from the onset of their formation during deglacial warming to the present (Fig. 3). Calculation of radiative forcing due to perturbations to the atmospheric carbon dioxide and methane budgets requires estimates of flux trajectories, which are a function of thermokarst-lake basin extent and flux rates. The extent of thermokarst basins across the landscape, which increased from the onset of deglaciation to the present, was assessed through evaluation of initiation or basal dates. Flux rates, which probably varied throughout the Holocene, were evaluated on the basis of time since basin initiation. Fluxes include methane and carbon dioxide emissions and the uptake of atmospheric carbon dioxide, resulting in carbon accumulation in lake sediments and peat. Methane emissions originated from two sources: thaw and decay of organic matter in yedoma permafrost and decomposition of younger organic matter fixed by photosynthesis in and around basins contemporary with thermokarst-lake cycling during deglaciation and the Holocene. Detailed methods, assumptions, and uncertainty assessment are provided in Supplementary Information section 1.6.

Modern lake sediments and mosses. Modern ecosystems provide a valuable, mechanistic understanding of alas deposit formation including their high carbon accumulation, high nutrient concentrations, and thriving benthic mosses in the lakes. From eight present-day yedoma lakes near Cherskii we extracted 21 sediment cores (49–210 cm) in 6.6-cm-diameter polycarbonate tubes using a piston hammer corer (Aquatic Instruments). Cores were extracted in the laboratory, visually described, and subsampled for quantification of dry bulk density and organic carbon and nitrogen. Terrestrial plant macrofossils from a subset of cores were radiocarbon dated (Supplementary Table 3). Modern benthic brown mosses sampled from present-day lakes were oven dried at 65 °C, ground through a Wiley Mill grinder (30 mesh), and redried at 65 °C before determination of carbon and nitrogen concentrations

on a LECO Truspec CHN analyser. Phosphorus was measured on a 300XL ICP following digestion in a solution of nitric acid and perchloric acid (Extended Data Table 5).

In the mineral-dominated sediments of lake cores (and alas permafrost exposures) we frequently observed nonapatite inorganic phosphorus macrofossils in the form of vivianite; however, we did not quantify vivianite concentrations.

Limnology. To compare the physiochemical characteristics of yedoma thermokarst lakes versus other types of arctic lakes we sampled yedoma and non-yedoma thermokarst lakes in North Siberia and compared results to non-yedoma lakes in Alaska (ref. 60 and G. Kling, personal communication, 10 April 2013) and the pan-Arctic[61]. During June, July and August 2002–2003 we sampled nine thermokarst lakes formed in Pleistocene-aged yedoma permafrost near Cherskii and 13 thermokarst lakes formed in Holocene-aged deposits on the adjacent Kolyma River floodplain. Three yedoma lakes were sampled biweekly during summer. Owing to the limited depth of Holocene ice wedges, floodplain thermokarst lakes were shallow (1–2.5 m). Yedoma lakes, formed by melt of deep, Pleistocene ice wedges and high intrasedimentary ground ice content, ranged in depth from 6 m to 17 m.

At the deepest known location in each lake, we collected water column physiochemical parameters along a water-column depth profile. We measured temperature, dissolved oxygen, pH, redox, salinity and specific conductivity using a handheld Quanta Hydrolab Multiparameter Sonde (Hach Hydromet). Bulk water samples were collected from various depths, filtered through Whatman GF/C glass microfibre filters, and analysed for orthophosphate (PO_4^{3-}, also expressed as soluble reactive phosphorus, SRP), nitrate (NO_3^-) and ammonium (NH_4^+) on a colorimetric autoanalyser (Astoria-Pacific) at the University of Florida. Ions were analysed using Astoria-Pacific method 305-A023-A00 for low-concentration ammonium, method 305-A173-A01 for low-concentration nitrite–nitrate, and method 305-A203-A00 for low-concentration orthophosphate. We measured dissolved organic carbon (DOC) at the University of Minnesota via high temperature combustion (Shimadzu). Chlorophyll-*a* was measured throughout the lake water column with a SCUFA Submersible Fluorometer (Turner Designs). We measured underwater photosynthetic active radiation (PAR, 400–700 nm) using an underwater spherical quantum sensor (LI-COR LI-193; LI-COR Biosciences) along the depth profiles to quantum irradiance following Wetzel and Likens[62]. The two-sided Mann–Whitney test was used to test differences in physiochemical parameters between the bottom water of North Siberian yedoma and floodplain lakes. Statistical analyses were performed in R[63].

Surface water temperature near the centres of five tundra yedoma lakes was measured in mid-August 2009 for comparison to boreal yedoma lakes. The Alaska lake data, collected in July and August 1997 and July 2009, were made available ref. 60 and G. Kling (personal communication, 10 April 2013). Relatively high nitrogen and phosphorus concentrations in the boreal yedoma thermokarst lakes (Extended Data Figs 3 and 4) are attributed to the release of nutrients from thawing yedoma surrounding lakes. This hypothesis is supported by soil and vegetation analyses in yedoma and non-yedoma surface types (Supplementary Information section 1.8).

31. Walter Anthony, K. M. & Anthony, P. Constraining spatial variability of methane ebullition in thermokarst lakes using point-process models. *J. Geophys. Res.* **118**, http://dx.doi.org/doi:10.1002/jgrg.20087 (2013).

32. Czudek, T. & Demek, J. Thermokarst in Siberia and its influence on the development of lowland relief. *Quat. Res.* **1**, 103–120 (1970).

33. Soloviev, P. A. Thermokarst phenomena and landforms due to frost heaving in Central Yakutia. *Biuletyn Peryglacjalny* **23**, 135–155 (1973).

34. Brouchkov, A., Fukuda, M., Fedorov, A., Konstantinov, P. & Iwahana, G. Thermokarst as a short-term permafrost disturbance, Central Yakutia. *Permafrost Periglacial Process.* **15**, 81–87 (2004).

35. Ferland, M. E., del Giorgio, P. A., Teodoru, C. R. & Prairie, Y. T. Long-term C accumulation and total C stocks in boreal lakes in northern Québec. *Glob. Biogeochem. Cycles* **26**, GB0E04 (2012).

36. Engstrom, D. R. & Rose, N. L. A whole-basin, mass balance approach to paleolimnology. *J. Paleolimnol.* **49**, 333–347 (2013).

37. Anderson, N. J., Dietz, R. D. & Engstrom, D. R. Land-use change, not climate, controls organic carbon burial in lakes. *Proc. R. Soc. Lond. B* **280**, 20131278 (2013).

38. Dean, W. E. Determination of carbonate and organic matter in calcareous sediments and sedimentary rocks by loss on ignition: comparison with other methods. *J. Sedim. Petrol.* **44**, 242–248 (1974).

39. Bundy, L. G. & Bremner, J. M. A simple titrimetric method for determination of inorganic carbon in soils. *Soil Sci. Soc. Am. J.* **36**, 273–275 (1972).

40. Reimer, P. J. et al. IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* **51**, 1111–1150 (2009).

41. Romanovskii, N. N. *Fundamentals of Cryogenesis of Lithosphere* 296–313 (Moscow Univ. Press, 1993).

42. Jorgenson, M. T. et al. Permafrost characteristics of Alaska. *Proc. Ninth Intl Conf. Permafrost* **3**, 121–122 (2008).

43. Grosse, G., Schirrmeister, L., Kunitsky, V. V. & Hubberten, H. W. The use of CORONA images in remote sensing of periglacial geomorphology: an illustration from the NE Siberian Coast. *Permafrost Periglacial Process.* **16**, 163–172 (2005).

44. Grosse, G., Schirrmeister, L. & Malthus, T. J. Application of Landsat-7 satellite data and a DEM for the quantification of thermokarst-affected terrain types in the periglacial Lena-Anabar coastal lowland. *Polar Res.* **25,** 51–67 (2006).
45. Veremeeva, A. & Gubin, S. Modern tundra landscapes of the Kolyma Lowland and their evolution in the Holocene. *Permafrost Periglacial Process.* **20,** 399–406 (2009).
46. Morgenstern, A., Grosse, G., Günther, F., Fedorova, I. & Schirrmeister, L. Spatial analyses of thermokarst lakes and basins in Yedoma landscapes of the Lena Delta. *Cryosphere* **5,** 849–867 (2011).
47. Morgenstern, A. Thermokarst and Thermal Erosion: Degradation of Siberian Ice-rich Permafrost. http://opus.kobv.de/ubp/volltexte/2012/6207/, PhD thesis, Potsdam Univ. (2012).
48. Grosse, G. *et al.* Distribution of late Pleistocene ice-rich syngenetic permafrost of the Yedoma Suite in East and Central Siberia, Russia. *USGS Open-file Rep.* 2013-1078 (2013).
49. Walker, D. A. *et al.* The Circumpolar Arctic vegetation map. *J. Veg. Sci.* **16,** 267–282 (2005).
50. Vtyurin, B. I. *Underground Ices of the USSR* [in Russian] 1–212 (Science, 1975).
51. Péwé, T. L., Journaux, A. & Stuckenrath, R. Radiocarbon dates and late-Quaternary stratigraphy from Mamontova Gora, unglaciated central Yakutia, Siberia, U.S.S.R. *Quat. Res.* **8,** 51–63 (1977).
52. Sher, A. V. *et al. Late Cenozoic of the Kolyma Lowland: XIV Pacific Science Congress, Tour Guide XI (Khabarovsk August 1979)* 1–116 (Academy of Sciences of the USSR, 1979).
53. Tomirdiaro, S. V. *Loess-ice Formation of Eastern Siberia in the Late Pleistocene and Holocene* 1–184 (Nauka, 1980).
54. Vasil'cuk, Y. K. *Oxygen Isotope Composition of Ground Ice Application to Paleogeocryological Reconstructions* [in Russian] Vols 1 and 2 (Russian Academy of Sciences and Lomonosov's Moscow University Publications, 1992).
55. Nikolaev, V. I., Mikhalev, D. V., Romanenko, F. A. & Brilli, M. Reconstruction of the conditions for North-East Russia permafrost formation on the isotope study results of Kolyma lowland key sections. [in Russian] *Ice Snow* **4,** 79–90 (2010).
56. Kanevskiy, M., Shur, Y., Fortier, D., Jorgenson, M. T. & Stephani, E. Cryostratigraphy of late Pleistocene syngenetic permafrost (yedoma) in northern Alaska, Itkillik River exposure. *Quat. Res.* (2011).
57. Schirrmeister, L., Siegert, C., Kunitzky, V. V., Grootes, P. M. & Erlenkeuser, H. Late Quaternary ice-rich permafrost sequences as a paleoenvironmental archive for the Laptev Sea Region in northern Siberia. *Int. J. Earth Sci.* **91,** 154–167 (2002).
58. Schirrmeister, L. *et al.* Periglacial landscape evolution and environmental changes of Arctic lowland areas for the last 60,000 years (western Laptev Sea coast, Cape Mamontov Klyk). *Polar Res.* **27,** 249–272 (2008).
59. Andreev, A. A. *et al.* Weichselian and Holocene palaeoenvironmental history of the Bol'shoy Lyakhovsky Island, New Siberian Archipelago, Arctic Siberia. *Boreas* **38,** 72–110 (2009).
60. Giblin, A., Luecke, C. & Kling, G. *Physical and Chemical Data for Various Lakes near Toolik Research Station, Arctic LTER Summer 2009* Arctic Long-Term Ecological Research Database, http://dx.doi.org/10.6073/pasta/1b77f4c8d8cc250ce0f90bbb17d9c976 (2009).
61. Lyons, W. B. & Finlay, J. C. in *Polar Lakes and Rivers: Limnology of Arctic and Antarctic Aquatic Ecosystems* (eds Vincent, W. F. & Laybourn-Parry, J.) 137–156 (Oxford Univ. Press, 2008).
62. Wetzel, R. G. & Likens, G. E. *Limnological Analyses* 3rd edn, 1–429 (Springer, 2000).
63. R Development Core Team. *R: A Language and Environment for Statistical Computing* http://www.R-project.org (R Foundation for Statistical Computing, 2009).
64. Kholodov, A. L. *et al.* Estimation of the organic carbon input to the arctic ocean due to erosion of Laptev and East-Siberian seashore. *Earth Cryosphere* **7,** 3–12 (2003).
65. Schirrmeister, L. *et al.* Periglacial landscape evolution and environmental changes of Arctic lowland areas for the last 60,000 years (western Laptev Sea coast, Cape Mamontov Klyk). *Polar Res.* **27,** 249–272 (2008).
66. Schneider, J., Grosse, G. & Wagner, D. Land cover classification of tundra environments in the Arctic Lena Delta based on Landsat 7 ETM+ data and its application for upscaling of methane emissions. *Remote Sens. Environ.* **113,** 380–391 (2009).
67. Hopkins, D. M. & Kidd, J. G. in *Proc. Fifth Intl Conf. Permafrost* (ed. Senneset, K.) 790–795 (Academic, 1988).
68. Jones, M. C., Grosse, G., Jones, B. M. & Walter Anthony, K. Peat accumulation in drained thermokarst lake basins in continuous, ice-rich permafrost, northern Seward Peninsula, Alaska. *J. Geophys. Res.* **117,** G00M07 (2012).
69. Kessler, M. A., Plug, L. J. & Walter Anthony, K. M. Simulating the decadal- to millennial-scale dynamics of morphology and sequestered carbon mobilization of two thermokarst lakes in NW Alaska. *J. Geophys. Res.* **117,** G00M06 (2012).
70. Ramaswamy, V. *et al.* in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (eds Houghton, J. T. *et al.*) 350–416 (Cambridge Univ. Press, 2001).
71. Joos, F. *et al.* An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake. *Tellus* **48,** 397–417 (1996).
72. Prather, M. *et al.* in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (eds Houghton, J. T. *et al.*) 239–287 (Cambridge Univ. Press, 2001).
73. Walker, D. A. & Everett, K. R. Loess ecosystems of northern Alaska: regional gradient and toposequence at Prudhoe Bay. *Ecol. Monogr.* **61,** 437–464 (1991).
74. Weintraub, M. N. in *Phosphorus in Action, Soil Biology* (eds Bunemann, E. K. *et al.*) Ch. 12, 295–316 (Springer, 2011).
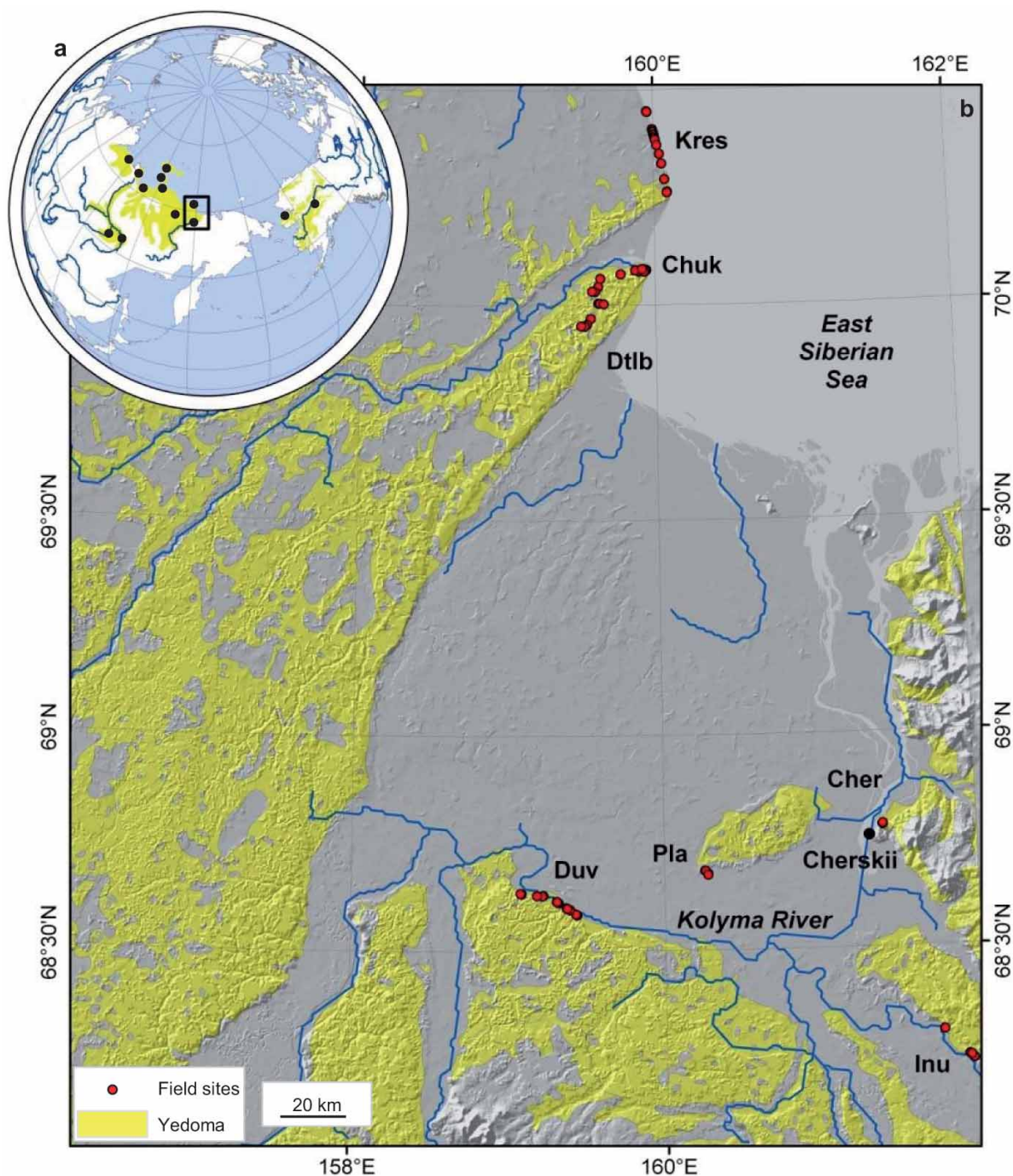
Extended Data Figure 1 | Map of main distribution of yedoma in the Beringia region in Siberia and Alaska (yellow regions). a, The Kolyma Lowland, considered largely covered by yedoma during the Last Glacial Maximum, now has only discontinuous yedoma coverage (yellow regions in **b**) owing to widespread destructive thermokarst and fluvial processes shaping the yedoma landscape since the early Holocene* (Supplementary Information section 1.1). Red dots in **b** indicate the locations of permafrost exposures sampled in boreal regions—Anuiy (Inu), Duvanii Yar (Duv), Plakhanski Yar (Pla), Cherskii (Cher)—and tundra regions—Chukochi Cape (Chuk and Dtlb) and Krestovskiy Cape (Kres). Literature data were synthesized from other western and eastern yedoma regions in Siberia[5,10,32,50–55,57–59,64–66] and Alaska[4,56,67–69], respectively (black dots in **a**). For map clarity, abundant lakes in the study regions were not plotted. **b**, Our central Beringia study region in the Kolyma Lowland in Northeast Siberia (small black frame in **a**; 60,000 km²) is characterized by yedoma hills, deep yedoma lake basins, and fluvial flood plains of the Kolyma River and its tributaries.
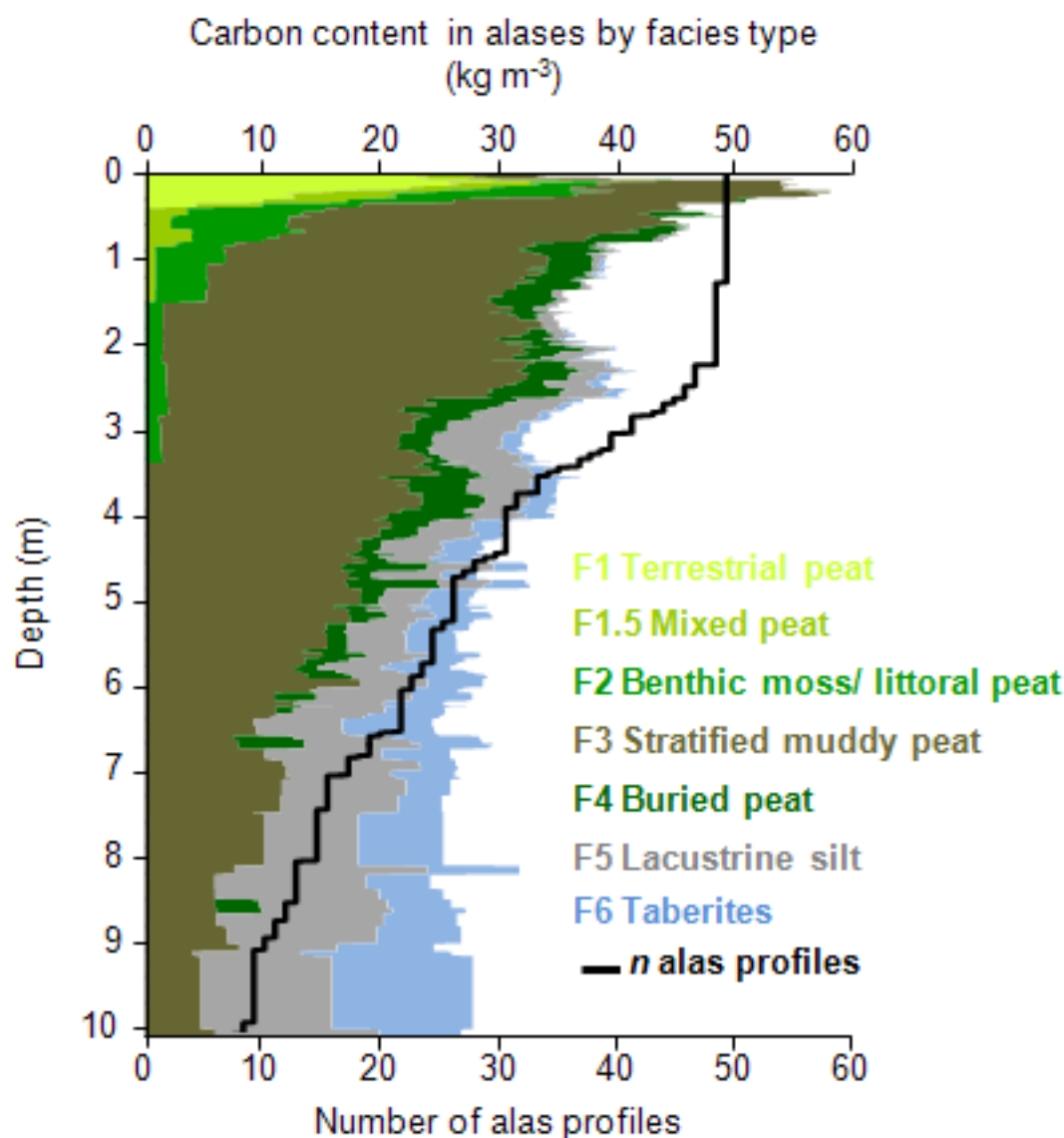
Extended Data Figure 2 | Relative contributions of facies F1–F6 to the average organic carbon content within the surface 10 m of North Siberian alases. The black line indicates the number of exposure profiles included in the observations. Extrapolating the Holocene* organic carbon component observed in these profiles to the extent of deep thermokarst basins in the yedoma region of Beringia (925,400 km², Supplementary Information section 1.6.1), we estimate the following Holocene* carbon pool sizes in the alases: 12 ± 2.5 Pg for 0–0.3 m, 36 ± 4.1 Pg for 0–1 m, 64 ± 4.3 Pg for 0–2 m, 89 ± 6.6 Pg for 0–3 m, 126 ± 9.0 Pg for 0–5 m, 144 ± 10.1 Pg for 0–7 m and 155 ± 11.6 Pg for 0–10 m. Error terms represent standard error at the 95% confidence limits derived by propagating uncertainties of the estimates of mean organic carbon bulk density for each depth interval, based on the interval size and number of field samples measured; additional uncertainty associated with the yedoma region extent is shown in Extended Data Table 3. Below 10 m, extrapolating our observation of Holocene* carbon in 7% of exposures, we estimate an additional 5 Pg C. Pleistocene carbon, also observed in the profiles and included in this figure, is accounted for in the regional-scale carbon mass balance calculation since these deposits extended deeper than we were able to expose in cross-section (Methods).
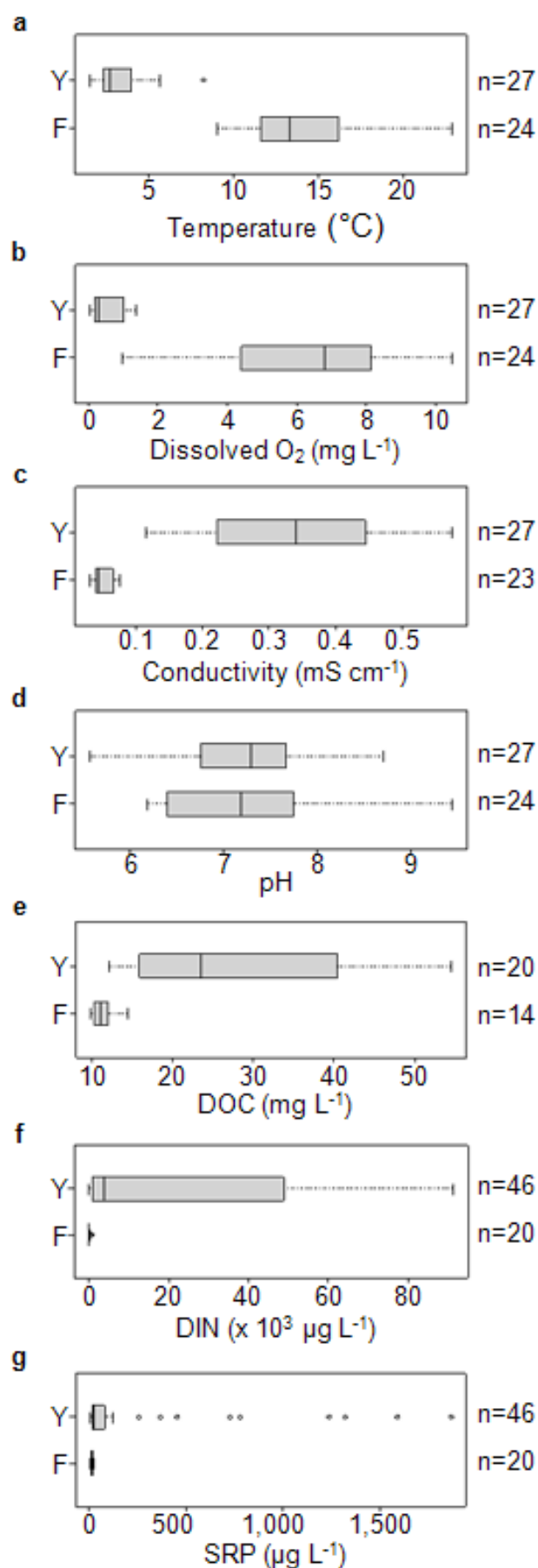
Extended Data Figure 3 | Box plots showing physiochemical characteristics of lake bottom water in thermokarst lakes formed in Pleistocene yedoma (Y) and non-yedoma Holocene floodplain (F) permafrost in the same region of North Siberia. DOC, dissolved organic carbon; DIN, dissolved inorganic nitrogen, dominated by ammonium; SRP, soluble reactive phosphorus. 'Conductivity' means specific conductivity. The number of samples n represent single observations per lake per day on different dates during June, July and August 2002–2003 from 9 yedoma and 13 floodplain thermokarst lakes (see Methods). The two-sample, two-sided Mann–Whitney test revealed differences between Y and F for all parameters except pH ($P < 0.01$).
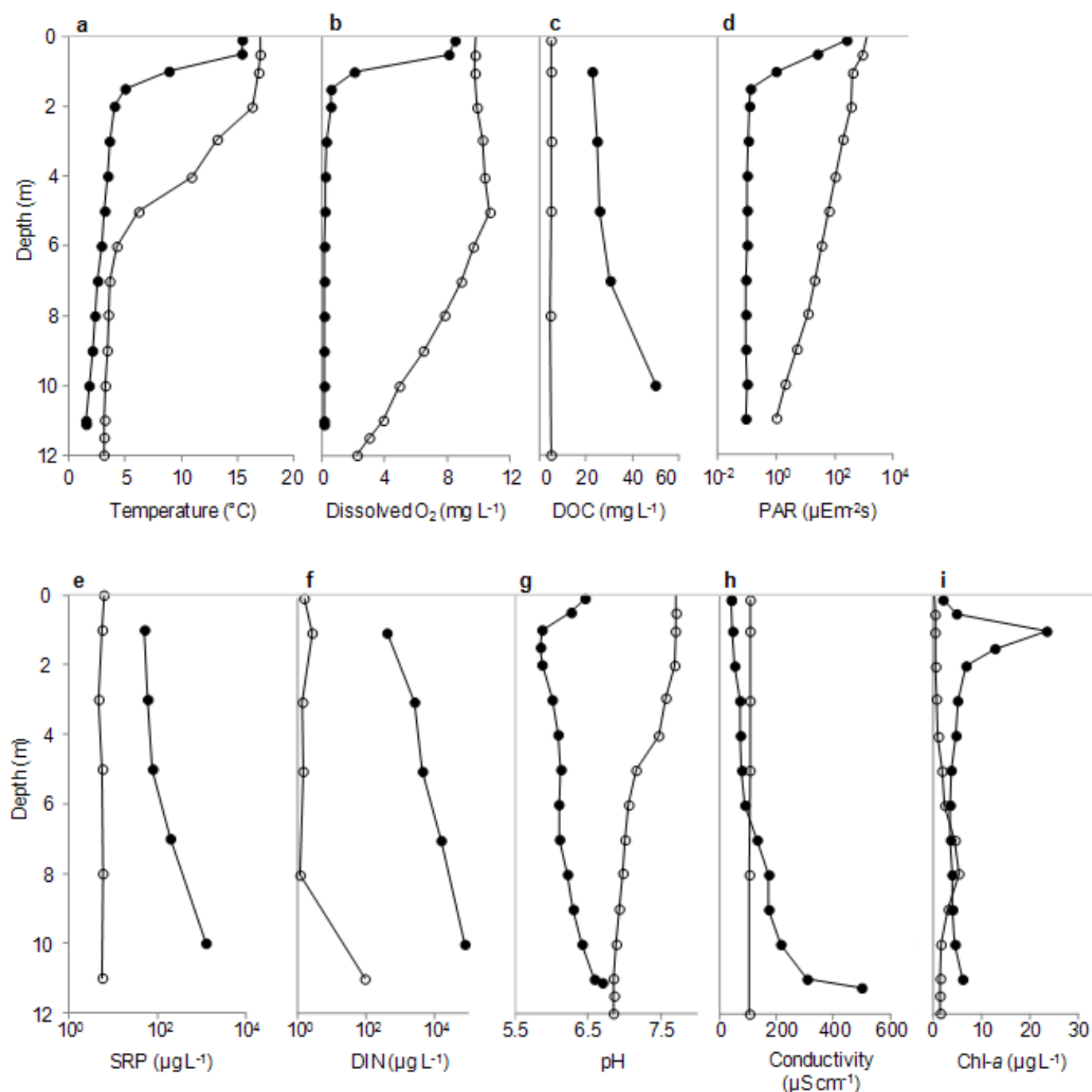
Extended Data Figure 4 | Comparison of physical, chemical, and biological characteristics of a yedoma lake and a non-yedoma lake of similar depth, volume and latitude in midsummer. Closed circles indicate the yedoma lake, Grass Lake (68.75° N, 161.38° W), near Cherskii, Russia. Open circles indicate the non-yedoma lake (68.64 °N, −149.61 °W) near Toolik Field Station, Alaska, USA. Both lakes were thermally stratified, but the yedoma lake had an anaerobic hypolimnion with exceedingly high concentrations of DOC, SRP, DIN and other solutes (indicated by specific conductivity). In addition, the yedoma lake had a much lower light environment, a colder lake bottom temperature, lower pH, and relatively high concentrations of chlorophyll-*a* in the epilimnion and dissolved ions in the hypolimnion. The Toolik Field Station data were from ref. 60 and G. Kling (personal communication, 10 April 2013).
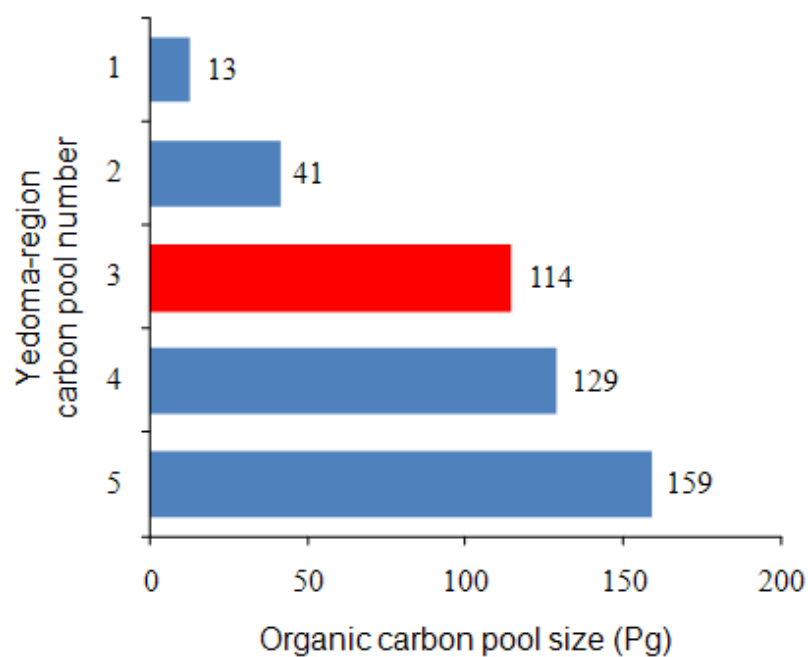
Extended Data Figure 5 | Organic carbon pools in the yedoma region. Our yedoma-region total organic carbon pool-size estimate (456 ± 45 Pg; Extended Data Table 3) is the sum of the following subset pools: (1) Holocene peat located above undisturbed yedoma permafrost; (2) yedoma that thawed, was reworked, and is now stored in thermokarst basins in facies F3 and F5; (3) taberite sediments representing *in situ* thawed, diagenetically altered yedoma in facies F6; (4) undisturbed yedoma in facies F7; (5) non-yedoma, Holocene* carbon stored in thermokarst basins in facies F1–F5 that was fixed via photosynthesis in and around the basins. Taberite deposits (red bar) are an important component of the yedoma-region total carbon pool that were not included in the recent yedoma-region carbon inventory of ref. 11.

Extended Data Table 1 | Physical and chemical characteristics of facies in North Siberian permafrost exposures

Facies	Moisture content (%) mean ± s.e., n	Dry bulk density (g soil cm⁻³) mean ± s.e., n	Inorganic C content (%) mean ± s.e., n	Organic matter content (%) mean ± s.e., n	Thicknessᵃ (m) mean ± s.e., n	Thicknessᵇ (m) mean ± s.e., median, n	Age (Cal. Yr. BP) mean ± s.e., n	median	(min to max)
F1	37 ± 4%, 14	0.21 ± 0.05, 14	0.96 ± 0.37%, 14	72 ± 4%, 21	0.13 ± 0.02, 49	0.17 ± 0.02, 0.15, 33	2,191 ± 1,770, n=4	642	(-2 to 7,479)
F1.5	39 ± 5%, 14	0.17 ± 0.02, 14	0.67 ± 0.13%, 14	70 ± 5%, 14	0.10 ± 0.04, 49	0.57 ± 0.13, 0.50, 9	3,734 ± 1,133, n=5	4,017	(699 to 6,523)
F2	82 ± 3%, 32	0.13 ± 0.02, 32	0.55 ± 0.16%, 32	82 ± 3%, 34	0.22 ± 0.08, 49	0.73 ± 0.21, 0.40, 15	4,368 ± 1,168, n= 9	3,533	(492 to 9,492)
F3	35 ± 1%, 132	1.05 ± 0.04, 132	0.97 ± 0.11%, 132	7.6 ± 0.5%, 249	3.11 ± 0.32, 49	3.62 ± 0.49, 2.46, 28	7,932 ± 693, n=23	8,613	(2,547 to 13,504)
F4	67 ± 4%, 20	0.39 ± 0.06, 40	0.92 ± 0.11%, 40	50 ± 4%, 67	0.20 ± 0.05, 49	0.24 ± 0.09, 0.05, 23	5,826 ± 923, n=16	5,038	(1,120 to 13,808)
F5	35 ± 2%, 61	1.09 ± 0.05, 61	0.83 ± 0.07%, 61	4.9 ± 0.2%, 80	1.12 ± 0.16, 49	1.62 ± 0.26, 0.80, 22	19,331 ± 7,318, n=5	15,331	(7,220 to 47,757)
F6	26 ± 1%, 14	1.35 ± 0.05, 14	1.01 ± 0.13%, 14	3.3 ± 0.1%, 40	-	-			
F7	31 ± 1%, 25	1.22 ± 0.05, 25	0.97 ± 0.07%, 25	4.4 ± 0.3%, 44	-	38.4 ± 1.8, 17†	n=4	>47,000	(42,319 to >47,000)

Shown are the facies gravimetric water content (percentage moisture), dry bulk density, inorganic carbon and organic matter contents, thickness, and maximum ¹⁴C ages. We collected and analysed 595 samples among the permafrost exposures in alases and yedoma. At several exposure sites, the average of field triplicate samples of yedoma (F7) were used as n = 1 in calculation of regional-scale means and standard error in this table and Extended Data Table 2. The purpose of this table is to show the physical and chemical characteristics of the facies; the mean values presented here were not used in upscaling calculations of Holocene* carbon stocks (see Methods and Fig. 2e). Values in this table were used together with other literature data synthesis in mass balance calculations to determine Pleistocene carbon stocks, given that we did not reach the base of the Pleistocene-aged deposits, F6 and F7, by digging in our exposures (see Methods). Facies thickness is expressed as the mean and standard error among all 49 studied exposures, including thicknesses with a zero value for instances where facies were not observed (Thicknessᵃ). Thicknessᵇ is the mean and median thickness of each facies among exposures where the facies was observed. Macrofossil ages represent the average maximum ages of the facies (not the average age of facies) because we often sampled the lower sections of the facies to date events such as lake formation and drainage. The large variability in maximum ¹⁴C ages within facies reflects lake formation at different times since the onset of the last deglaciation about 14 kyr ago; these ¹⁴C data are used in this study as secondary independent evidence for Pleistocene versus Holocene* carbon fractions in facies (Supplementary Information section 1.4).
† Undisturbed yedoma thickness determined for a larger sample size from map-based analysis. Measured yedoma thicknesses at studied exposures was 31.5 ± 5.9 m (n = 4). Our carbon stock and pool size calculations (Extended Data Table 3) assume an average undisturbed yedoma deposit thickness of 25 ± 5 m, as determined from extensive literature[10,32,50–56].

Extended Data Table 2 | Organic carbon concentrations in yedoma (F7) and taberites (F6) and organic carbon bulk density from various subregions in North Siberia

| Subregion | (a) Organic carbon content (%) | | | | | | (b) Yedoma (F7) organic carbon bulk density (%) |
| | Yedoma (F7) | | | Taberites (F6) | | | |
	Mean	s.e.	n	Mean	s.e.	n	Difference	(kg C m^{-3})
Chukochi Cape*	2.41	± 0.22	17	1.76	± 0.07	19	27%	27.88
Anuiy*	2.96	± 0.74	7	2.16	± 0.22	4	27%	30.75
Duvanni Yar*†	2.77	± 0.39	8	1.98	± 0.10	10	28%	24.62
Laptev Sea†	3.38	± 0.17	468	2.70	± 1.40	9	20%	
East Siberian/Laptev Seas‡	2.03	± 0.13	24	1.29	± 0.07	29	36%	
Plakhanski Yar*								26.66
Cape Mamontov Klyk†								24.84
Lena Delta†								30.32
Bykovsky Peninsula†								25.04
Bol'shoy Lyakhovsky Island†								17.72
Mean ± SE							28 ± 12%	25.98 ± 1.45

* Data collected by this study; †data from ref. 5; ‡data from ref. 64. **a**, The organic matter content of yedoma was significantly higher than that of taberites (two-way analysis of variance, ANOVA, $F = 8.12_{22,1}$, $P = 0.0061$) at the $\alpha = 0.05$ level. Interaction effects of facies by zone (boreal versus tundra) or facies by subregions listed in the table were not significant. This indicates that the organic matter contents of yedoma and taberites are statistically distinct irrespective of the subregion or zone in which they are measured. Across subregions, the organic matter content of yedoma was on average 28% higher than that of taberites (4%–52%, lower and upper 95% confidence interval). The large uncertainty in the mean difference between yedoma and taberite organic carbon concentration is largely due to unpaired sampling in the literature data sets († and ‡), which had different goals in their study designs. The more tightly constrained difference in our field data sets results from pair-wise sampling of adjacent yedoma/alas systems in our study, which was designed specifically to assess site-specific carbon loss. **b**, The yedoma organic carbon bulk density was calculated from field measurements in eight subregions of North Siberia. The value used for Duvanni Yar was the average of both independent studies. Values in this table do not account for ice-wedge volume.

Extended Data Table 3 | Calculations (a) and uncertainty analysis (b) of estimated carbon pool sizes and fluxes in the yedoma region

a

Land cover class	Fraction	Area (10^6 km²)	Cumulative loss to atmosphere (kg C_P m⁻²)	Cumulative loss to atmosphere (Pg C_P)	Export to sea (kg C_P m⁻²)	Export to sea (Pg C_P)
Total yedoma region	100%	1.322 ± 0.132†		100 ± 34‖		45 ± 12**
Undisturbed yedoma	30%	0.397 ± 0.040†				
Thermokarst lakes/ alases	55%	0.727 ± 0.073†	131 ± 45‡	95 ± 34¶		
Thermoerosional gullies	15%	0.198 ± 0.020†	27 ± 9§	5 ± 2#	227 ± 49☆	45 ± 12**

Land cover class	Present-day stocks (kg C_P m⁻²)	Present-day stocks (kg C_H m⁻²)	Regional soil carbon pools (Pg C_P)	Regional soil carbon pools (Pg $C_{H\text{-}tk}$)	Regional soil carbon pools (Pg C_H)	Regional soil carbon pools (Pg C_{total})
Total yedoma region			284 ± 40¶¶	159 ± 24¶¶	172 ± 19¶¶	456 ± 45¶¶
Undisturbed yedoma	325 ± 69††		129 ± 30		13 ± 1##	142 ± 30
Thermokarst lakes/ alases	194 ± 30‡‡	172 ± 19‖‖	141 ± 26	125 ± 19	125 ± 19	266 ± 32
Thermoerosional gullies	70 ± 9§§	172 ± 19‖‖	14 ± 2	34 ± 5	34 ± 5	48 ± 6

b

Uncertainty factor	Value used	Uncertainty	Loss to atmosphere (Pg C_P)	Export to sea (Pg C_P)	Regional soil carbon pools (Pg C_P)	Regional soil carbon pools (Pg $C_{H\text{-}tk}$)	Regional soil carbon pools (Pg C_{total})
Yedoma-region area	1.322 x 10^6 km² ± 10%	± 20%	100 ± 38	45 ± 16	284 ± 52	159 ± 36	456 ± 60
Pleistocene *vs*. Holocene* C attribution							
Fraction C_p decomposed (F3)	56 ± 24%	95 ± 0%	119 ± 33	-	265 ± 38	175 ± 26	453 ± 43
Fraction C_p decomposed (F5)	28 ± 12%	40 ± 0%	104 ± 30	-	280 ± 34	163 ± 24	456 ± 39
Thickness of F3 and F5	mean	median	96 ± 31	-	288 ± 41	-	460 ± 45
Holocene* carbon stocks in alases	mean	median	-	-	-	153 ± 23	450 ± 44
Woody biomass	2.3 kg C m⁻²	0.7 kg C m⁻² to	-	-	-	158 ± 24	455 ± 45
(forest succession & density)		5.1 kg C m⁻²				161 ± 24	458 ± 45
Woody biomass extent	0.889 x 10^6 km²	+47%	-	-	-	160 ± 24	457 ± 45
(boreal *vs*. boreal + tundra zones)							
Thermokarst-affected area	70%	-7% to	93 ± 32	42 ± 11	294 ± 43	148 ± 22	457 ± 47
(percent of region)		+11%	112 ± 38	50 ± 13	267 ± 37	177 ± 27	454 ± 43

We included present-day carbon stocks, regional soil carbon pools for different landscape cover classes of the yedoma region (undisturbed yedoma, thermokarst lakes and alases, and thermoerosional gullies) and cumulative carbon export to the atmosphere or sea. Pleistocene yedoma-derived carbon is indicated as C_P; Holocene* carbon is C_H. The regional soil carbon pool $C_{H\text{-}tk}$ specifies Holocene* carbon in thermokarst basins only, while C_H includes also the Holocene surface soil overlying undisturbed yedoma. Uncertainties are represented as propagated standard errors unless otherwise noted. In **b**, '-' indicates that an additional uncertainty had no impact on the calculation of a particular carbon pool. Uncertainty sources, conservative basis for assumptions, additional minor uncertainties, and impacts on upscaling results are discussed in Supplementary Information section 1.7.

† See Methods; the error term represents 10% uncertainty (Supplementary Information section 1.7.3).

‡ 131 ± 76 kg C_P m⁻² (lost to the atmosphere from thermokarst lakes as methane and carbon dioxide) = 325 ± 69 kg C_P m⁻² (yedoma carbon stock, see ††) minus 194 ± 30 kg C_P m⁻² (C_P remaining in lake basins, see ‡‡).

§ 27 ± 9 kg C_P m⁻² (lost to the atmosphere from thermoerosional gullies as methane and carbon dioxide) = 325 ± 69 kg C_P m⁻² (yedoma carbon stock, see ††) minus 70 ± 9 kg C_P m⁻² (C_P remaining in lake basins, see §§) minus 227 ± 49 kg C_P m⁻² (C_P exported to sea, see ☆); this assumes 28 ± 12% of the remaining thawed yedoma organic carbon decomposed under anaerobic conditions in F5 and F6, forming methane and carbon dioxide that escaped to the atmosphere (see Extended Data Table 2 and Supplementary Information section 1.4).

‖ 100 ± 56 Pg C_P (total Pleistocene yedoma carbon lost to the atmosphere as CH₄ and CO₂) = 95 ± 56 Pg C_P (loss from thermokarst lakes, see ¶) plus 5 ± 2 Pg C_P (loss from thermoerosional gullies, see #).

¶ 95 ± 56 Pg C_P (C_P lost to the atmosphere as methane and carbon dioxide from thermokarst lakes) = 131 ± 76 kg C_P m⁻² (see ‡) times 727,100 ± 72,700 km² (thermokarst lake/alas area, see †).

5 ± 2 Pg C_P (C_P lost to the atmosphere as methane and carbon dioxide from thermoerosional gullies) = 27 ± 9 kg C_P m⁻² (see §) times 198,300 ± 19,800 km² (thermoerosional gully area, see †).

☆ 227 ± 49 kg C_P m⁻² (thawed Pleistocene yedoma carbon exported to the sea, including river deltas and continental shelves) = 0.7 (eroded fraction; Supplementary Information section 1.5) times 325 ± 69 kg C_P m⁻² (yedoma carbon stock, see ††).

** 45 ± 12 Pg (C_P exported to sea) = 227 ± 49 kg C_P m⁻² (see ☆) times 198,300 ± 19,800 km² (thermoerosional gully area, see †).

†† 325 ± 69 kg C_P m⁻² (original, undisturbed yedoma carbon stock) = 25.98 ± 1.45 C_P m⁻³ (Extended Data Table 2) × 25 ± 5 m yedoma deposit thickness × 50% ± 3% ice wedges (Methods).

‡‡ 194 ± 30 kg C_P m⁻² (Pleistocene yedoma-derived carbon remaining in alases) = 29 ± 16 kg C_P m⁻² (in facies F3) + 22 ± 4 kg C_P m⁻² (in facies F5) + 143 ± 25 kg C_P m⁻² (in facies F6). Assumes 28% ± 12% of original yedoma organic carbon decomposed under anaerobic conditions in F5 and F6, and 56% ± 24% under aerobic and anaerobic conditions in F3, forming methane and carbon dioxide that escaped to the atmosphere (see Extended Data Table 2 and Supplementary Information section 1.4). Calculation of the mineral fraction thickness of F3 and F5, based on dry bulk density mixing ratios, is summarized in equations (1) and (2) (Supplementary Information section 1.4). We determined the mean thickness of F6 based on a conservation of mass of yedoma on the landscape, accounting for the difference in dry bulk density between frozen (F7) and thawed yedoma (F6).

§§ 70 ± 9 kg C_P m⁻² (Pleistocene yedoma-derived carbon remaining in thermoerosional gullies) = 22 ± 4 kg C_P m⁻² (in facies F5) + 48 ± 9 kg C_P m⁻² (in facies F6). Assumes 28% ± 12% of original yedoma organic carbon decomposed under anaerobic conditions in F5 and F6, and 56% ± 24% under aerobic and anaerobic conditions in F3, forming methane and carbon dioxide that escaped to the atmosphere (see Extended Data Table 2 and Supplementary Information section 1.4).

‖‖ 172 ± 19 kg C_H m⁻² is the area-weighted average (32% tundra, 68% boreal) of Holocene* organic carbon stocks in tundra alases (122 ± 16 kg C m⁻², n = 18 sites) and boreal alases (196 ± 27 kg C m⁻², n = 10 sites) (Fig. 2e).

¶¶ The regional soil carbon pools are the sum of the Pleistocene yedoma-derived and/or Holocene* carbon pools in each of the land cover classes (undisturbed yedoma, thermokarst lakes/alases, and thermoerosional gullies). The carbon pools within each land cover class were calculated as the product of the associated carbon stocks and cover class areas.

13 ± 1 kg C_H m⁻² is the surface soil (0–100 cm) Holocene organic carbon pool overlying upland yedoma estimated from the distribution of histels, orthels and turbels in the Northern Circumpolar Soil Carbon Database (NCSCD)[13] (Supplementary Information section 3.5). The error term is a minimum estimate, representing the 10% uncertainty in the yedoma region extent (this study, Supplementary Information section 1.7.3); errors associated with the NCSCD carbon stocks are unknown.

Extended Data Table 4 | Greenhouse gas parameters for atmospheric model

Gas	Index, i	A_i * (10^{-13} W m^{-2} kg^{-1})	f_i [†]	τ_i [‡] (yr)	ε_i *
CO_2	0	0.0198	0.176	*infinite* [§]	1
	1	0.0198	0.138	421	1
	2	0.0198	0.186	70.6	1
	3	0.0198	0.242	21.4	1
	4	0.0198	0.259	3.42	1
CH_4	5	1.3	1	12.0	1.3

See equation (3) and Supplementary Information section 1.6.6.
* Ref. 70.
† Ref. 71.
‡ CO_2 values from ref. 71; CH_4 value from ref. 72.
§ Modelled as 10^8 yr.

Extended Data Table 5 | Mean nitrogen (N) and phosphorus (P) concentrations in ice wedges, soils and present-day vegetation

Pool	Element	Yedoma upland		Floodplain	P-value
Ice wedges					
(ppm)	Total N	1.79		-	
(mg L^{-1})	Total dissolved P	0.29		-	
Soil soluble (g/m^2)	N	2.43	=	2.28	--
	P	1.27	>	0.18	***
	N:P	2.21	<	12.5	***
Soil bulk (g/m^2)	N	222	<	494	***
	P	39	<	63	***
	N:P	6	=	6	--
Salix foliage (%)	N	2.07	=	2	--
	P	0.3	<	0.42	**
	C:N	24	=	25	--
	N:P	7.3	>	5.8	*
Salix litter (%)	N	0.71	<	0.94	**
	P	0.21	>	0.09	***
	C:N	64	>	50	*
	N:P	3.79	<	10.82	***
	% N retranslocated	67	=	64	--
	% P retranslocated	30	<	80	***
	N:P retranslocated	2.2	>	0.8	***
Salix litter soluble (mg/g)	C	8.9	>	5.8	*
	N	0.83	>	0.71	*
	P	2.29	>	0.55	***
	C:N	115	>	90	*
	N:P	0.44	<	1.5	**
	% mass lost	26	>	15	*
Benthic brown moss (%)	N	1.95		-	
	P	0.12		-	
	N:P	20.37		-	

Results are shown for active layer soil and shrubs for two permafrost-dominated landscape positions near Cherskii in northeast Siberia (Pleistocene yedoma upland and Holocene floodplain), and mean N and P for Pleistocene-aged yedoma ice wedges and yedoma thermokarst-lake mosses. Means are for $n = 4$ terrestrial sites on the Holocene floodplain, $n = 6$ sites terrestrial sites in the Pleistocene yedoma upland, $n = 6$ yedoma ice wedges, and $n = 11$ mosses from eight yedoma thermokarst lakes. P values are from Mann–Whitney U tests. Significant differences at *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$. Results for *Salix* (willow) are used to illustrate contrasts in litter nutrient retention between landscapes because it was common in both environments. Phosphorus retranslocation was significantly lower for yedoma upland versus floodplain plants ($P < 0.01$), while N retranslocation was highest in the yedoma upland. Higher N than P retranslocation together with higher soil P availability suggests that P is available in excess of N on the yedoma upland. Bulk N and P concentrations of yedoma were similar to other arctic soils[73,74], but high soil P availability, low plant P retranslocation, and high aquatic P concentrations support the idea that yedoma soils have very high bioavailable P pools. High litter P concentration and high litter soluble P pools also indicate high P availability for leaching and runoff to lakes. Further evidence for high P mobility in yedoma soils is that the yedoma thermokarst-lake P concentrations (Extended Data Figs 3 and 4) are much higher than in other arctic lakes in Alaska and Canada[61]. Phosphorus and nitrogen concentrations in benthic mosses of yedoma thermokarst lakes (*Warnstorfia fluitans, W. exannulatus, D. anduncus, D. vernicosis, Scorpidium revolvens, S. scorpiodes, Calliergon giganteum*) were also well above critical concentrations (about 0.09% P, about 1% N)[20].

LETTER

Zircons reveal magma fluxes in the Earth's crust

Luca Caricchi[1], Guy Simpson[1] & Urs Schaltegger[1]

Magma fluxes regulate the planetary thermal budget, the growth of continents and the frequency and magnitude of volcanic eruptions, and play a part in the genesis and size of magmatic ore deposits[1–4]. However, because a large fraction of the magma produced on the Earth does not erupt at the surface[2,5], determinations of magma fluxes are rare and this compromises our ability to establish a link between global heat transfer and large-scale geological processes. Here we show that age distributions of zircons, a mineral often present in crustal magmatic rocks[6], in combination with thermal modelling, provide an accurate means of retrieving magma fluxes. The characteristics of zircon age populations vary significantly and systematically as a function of the flux and total volume of magma accumulated in the Earth's crust. Our approach produces results that are consistent with independent determinations of magma fluxes and volumes of magmatic systems. Analysis of existing age population data sets using our method suggests that porphyry-type deposits, plutons and large eruptions each require magma input over different timescales at different characteristic average fluxes. We anticipate that more extensive and complete magma flux data sets will serve to clarify the control that the global heat flux exerts on the frequency of geological events such as volcanic eruptions, and to determine the main factors controlling the distribution of resources on our planet.

Zircon is a common accessory mineral in magmatic rocks emplaced in continental crust, incorporating radioactive elements that enable the determination of its crystallization age[6]. High-precision single-zircon U–Pb ages in plutons (that is, in magmatic bodies cooled at depth) typically yield crystallization ages spanning 10^5–10^6 years (refs 7–11), which is clearly larger than analytical uncertainty and reproducibility. This level of analytical precision enables us to identify important differences between populations of zircon ages in terms of mode, median and standard deviation (Extended Data Fig. 1). These differences are directly controlled by the total number of zircons crystallizing in a certain magma volume over time and (because zircons crystallize within a restricted range of temperature) by the temporal evolution of temperature in a given magmatic
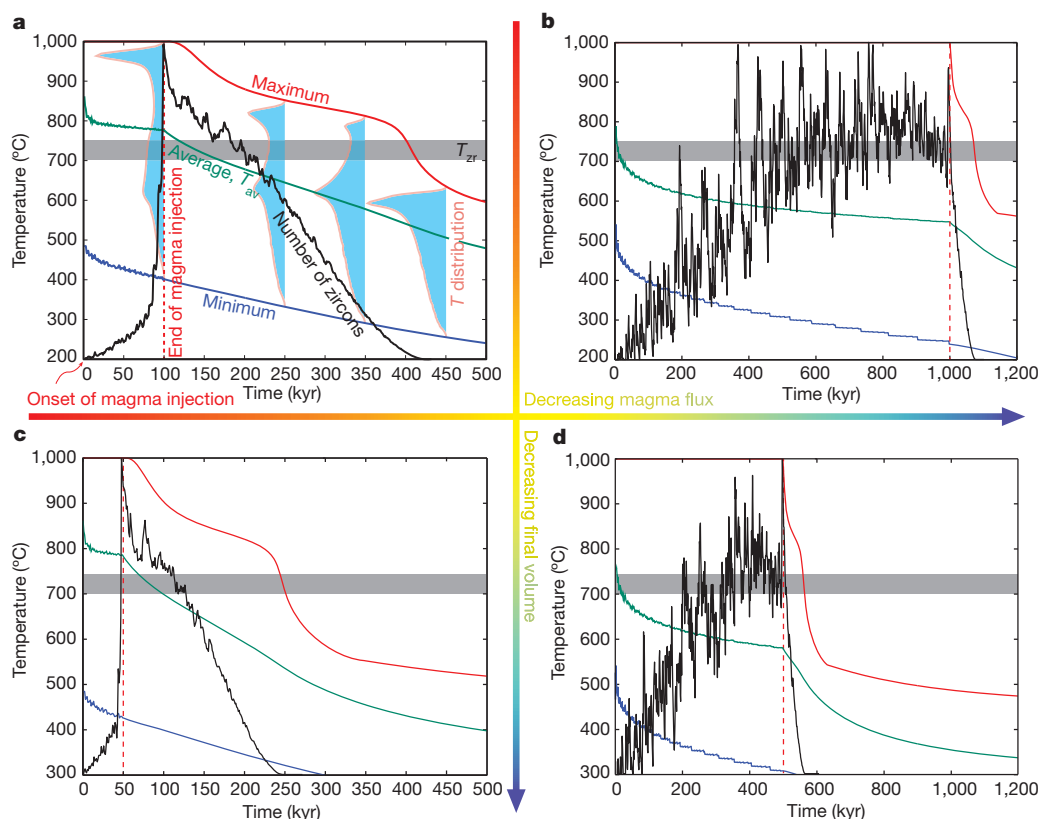
Figure 1 | Time evolution of maximum, average and minimum temperature, along with the number of newly crystallized zircons for different magma fluxes and final volume of accumulated magma. a, b, Final volume is 1,000 km³. c, d, Final volume is 500 km³. a, c, Magma flux is 0.01 km³ yr⁻¹. b, d, Magma flux is 0.001 km³ yr⁻¹. The number of new zircons that crystallize within the magma body (black line) is computed for each time slice of one kiloyear. The grey region indicates the interval of zircon crystallization temperature, T_{zr}. Red, green and blue curves provide the evolution over time of maximum, average and minimum temperature within the magma body, respectively. The light blue shapes in a show the distribution of temperature within the intrusion at different times.

[1]Section of Earth and Environmental Sciences, University of Geneva, rue des Maraîchers 13, CH-1205 Geneva, Switzerland.

system. In turn, the thermal evolution of a magmatic body depends on its shape and the modalities of heat release and is directly related to the flux of magma and to its volume[12,13]. Therefore, U–Pb ages of zircon crystals from magmatic rocks provide an integrated temporal record of the thermal history of such rocks[6–8].

To link the statistical parameters of zircon crystallization age spectra to the rate of magma injection and the volume of a magmatic system, we performed thermal modelling, systematically varying the final volume of injected magma between 30 km^3 and 10,000 km^3 and the magma flux between 10^{-4} km^3 yr^{-1} and 10^{-1} km^3 yr^{-1}. The flux of magma was kept constant because short-term fluctuations of the magma input rate, which are typical of magma transfer in the crust[3,7,8,14,15], do not appreciably affect the long-term evolution of temperature of magmatic systems[12] and therefore the fundamental characteristics of a population of zircon ages.

We modelled the injection of magma at depths of 10 km (Methods) into a crust with an initial geothermal gradient of 30 °C km^{-1}. We considered the intrusion of granodioritic magma at 1,000 °C. Composition and temperature were selected to simulate the intrusion of magmas characteristic for the upper continental crust at water-saturated or moderately water-undersaturated (by about 1 wt%) liquidus temperature[16]. Considering a granodioritic magma composition and initial zirconium concentrations between 50 parts per million (p.p.m.) and 250 p.p.m. (http://georoc.mpch-mainz.gwdg.de), we calculated a plausible range of zircon saturation temperatures[6,17] T_{zr} between 700 °C and 750 °C. To compute the number of new zircons crystallized within a magma body at any one time (that is, the zircon age spectra), we assume that one new zircon grows every kiloyear (kyr) for each unit volume of magma between 700 °C and 750 °C. We traced back in time the temperature evolution of each unit volume of magma and determined the total number of zircons crystallized within single parcels of magma by calculating the total time such magma volumes spent at T_{zr}. The zircon age population for the

entire magma body is then reconstructed by summing the number of zircons that crystallized in each unit magma volume (Fig. 1).

We do not consider zircon resorption and overgrowth. To test the sensitivity of the calculated zircon age populations to the various assumptions included in the model, we also performed calculations with a different geothermal gradient (50 °C km^{-1}), a broader T_{zr} range (700–800 °C; ref. 18), a lower temperature for the injected magma (900 °C) and various geometries of magma emplacement (Methods and Supplementary Table 1).

To illustrate how the thermal evolution of a magmatic body relates to zircon age spectra, we present four end-member scenarios calculated for different fluxes and total volumes of accumulated magma (Fig. 1). We discuss our data in terms of time, with zero corresponding to the onset of magma injection and therefore the oldest zircon age. We use the term 'crystallization time' to refer to the specific time after the onset of magma intrusion when a particular zircon or group of zircons grew. In all simulations the continuous injection of magma in the crust leads to a net decrease of average temperature T_{av} with time, independently of the rate of magma injection (Fig. 1). This occurs because with increasing volume of the magmatic reservoir, the surface over which heat is released also increases and the flux of magma into colder crust becomes insufficient to keep the intrusion temperature constant[19,20].

Figure 1 shows that the mode of the zircon population (recording the time at which the largest number of zircons crystallized) corresponds (or is close) to the time at the end of magma injection. Thus, for magma bodies of the same volume, a decrease of magma flux, which implies longer duration of the magma injection episode, leads to younger modal zircon ages (that is, longer times from the onset of magma injection; compare Fig. 1a with Fig. 1b, and Fig. 1c with Fig. 1d). The positive slope of the spectra on the older side of a population of zircon ages is inversely proportional to the time required to assemble a magma body of a given volume and, therefore, of the magma flux (Fig. 1). The negative slope
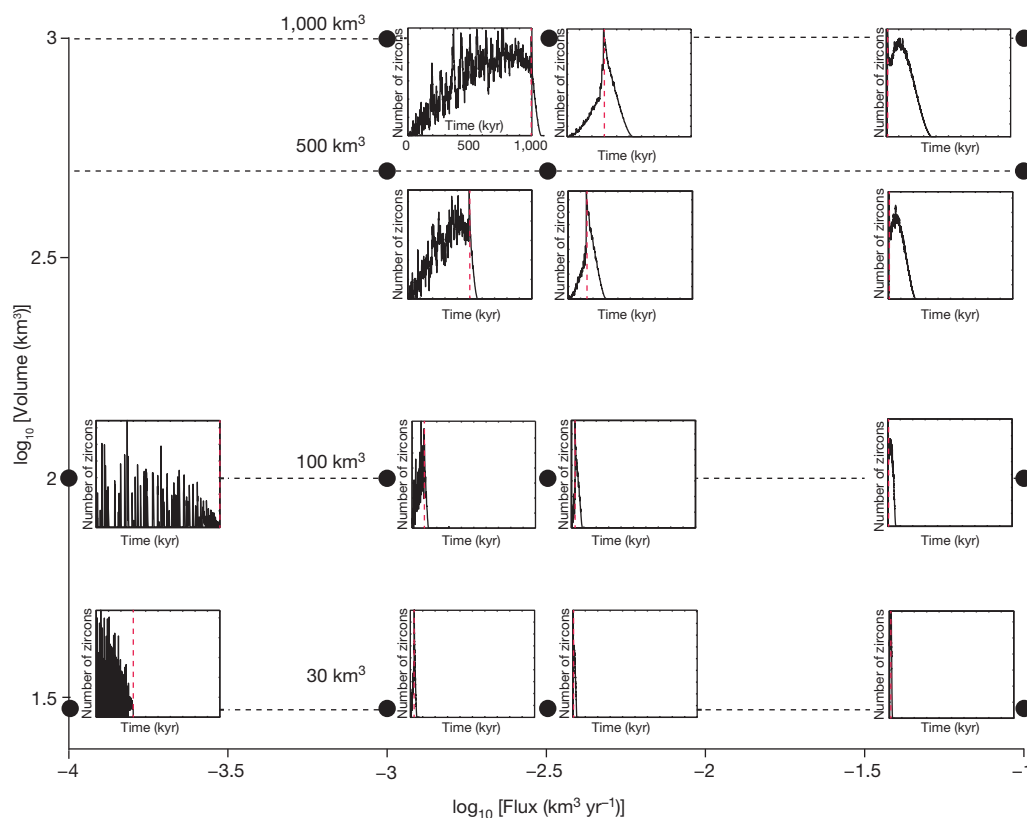
Figure 2 | Map of selected models performed in the magma flux versus final injected volume space. Black dots show the simulated magma fluxes and total injected volumes in logarithmic scale and time in kiloyears. Simulations performed for a total of ten million years are not shown. The final zircon populations calculated for the different runs are shown by the black curves in the boxes; the horizontal scale is the same (one million years) for all boxes. The red dashed lines mark the end of magma injection.

on the youngest side of the zircon age population is directly proportional to the rate at which the intrusion cools through the range of T_{zr} once magma injection stops. Hence, large bodies of magma emplaced over a relatively short time, which cool slower than smaller ones, are characterized by steep slopes on the old side of the age spectra, and gentler slopes on the young side (Fig. 1a, c).

The mode, median and standard deviation of the calculated populations of zircon crystallization times vary systematically as function of the final volume of the magma body and the average magma flux (Figs 2 and 3). An increase of magma flux at constant final volume and a decrease of final volume at constant magma flux both lead to a decrease of mode, median and standard deviation of the distribution of zircon crystallization times (Fig. 3). Slightly more complicated, but traceable, relationships are obtained for the mode of the populations at high magma fluxes and final volumes of the magma body lower than 100 km^3 (Fig. 3).

Skewness also varies consistently in our model but is not presented here because it requires a number of zircon analyses that exceeds current analytical capabilities to yield reliable estimates. Low magma fluxes produce negatively skewed distributions of crystallization times because zircon crystallization proceeds throughout the injection phase but terminates abruptly once injection ceases and the entire reservoir cools rapidly below T_{zr} (Figs 1 and 2). Conversely, when magma fluxes are high, the time required to accumulate a given volume of magma is short and the rate of cooling below T_{zr} is slower than in the first case, which produces positively skewed distributions of zircon crystallization times (Figs 1 and 2).

We have successfully inverted calculated zircon crystallization–time spectra and obtained precise (and unique) values of the magmatic flux and final volume of the magma body used in the simulations, thereby confirming the validity of this approach. Simulations performed with different parameters (such as different intrusion geometries, geothermal gradient, magma intrusion temperature and 700 °C < T_{zr} < 800 °C) result in uncertainties on the final estimates of magma flux and volume of injected magma obtained by our inversion method that are well within those related to the analytical error for natural zircon populations (Methods and Supplementary Table 1).

To assess whether porphyry deposits, super-eruptions and plutons potentially develop with characteristic magma fluxes and timescales, and to test the robustness of our approach, we applied our method to populations of zircon ages obtained for different systems[8,9,21,22] (Fig. 4). The ages were converted into crystallization times by subtracting from the oldest age those of each zircon. Zircons with an age difference on the older side larger than the 2σ uncertainty with respect to the oldest zircon were not considered part of the same population and were excluded. We binned crystallization times using the 2σ average uncertainty of the measured zircon ages and we calculated the resulting uncertainty on the mode, median and standard deviation of populations, applying the bootstrap method with a 95% confidence interval (Figs 4a–c).

We used data from two plutons: the Lago della Vacca complex (Italy) and the Torres del Paine laccolith (Patagonia; Fig. 4a, Extended Data Fig. 3a). The estimates obtained for these systems are in good agreement with previous independent determinations[7–10,15,23] and confirm that crustal intrusions, for which no evidence of associated volcanic activity exists, are emplaced in the crust at relatively low magmatic fluxes[1,2,24] (Fig. 4a, d) . The intrusion in the Bingham Canyon, USA, is the source of a world-class porphyry copper deposit and its volume has been demonstrated to be sufficient to provide the total amount of metals estimated to be present within the deposit[22]. Analysis of the zircon population from this system gives a total volume of injected magma that is in agreement with that estimated by ref. 22 (Fig. 4b). The investigation of a zircon data set from a smaller porphyry system (Coroccohuayco, Peru) shows that the associated intrusion has a smaller volume than does Bingham Canyon but that their magma fluxes and the durations of their magmatic episodes were comparable (Extended Data Figs 3b and 4b, d). These data support the idea that a favourable condition for the genesis of porphyry deposits is the prolonged intrusion of magma in the crust at relatively low fluxes[4] (Fig. 4d). The existence of a direct relationship between the volume of
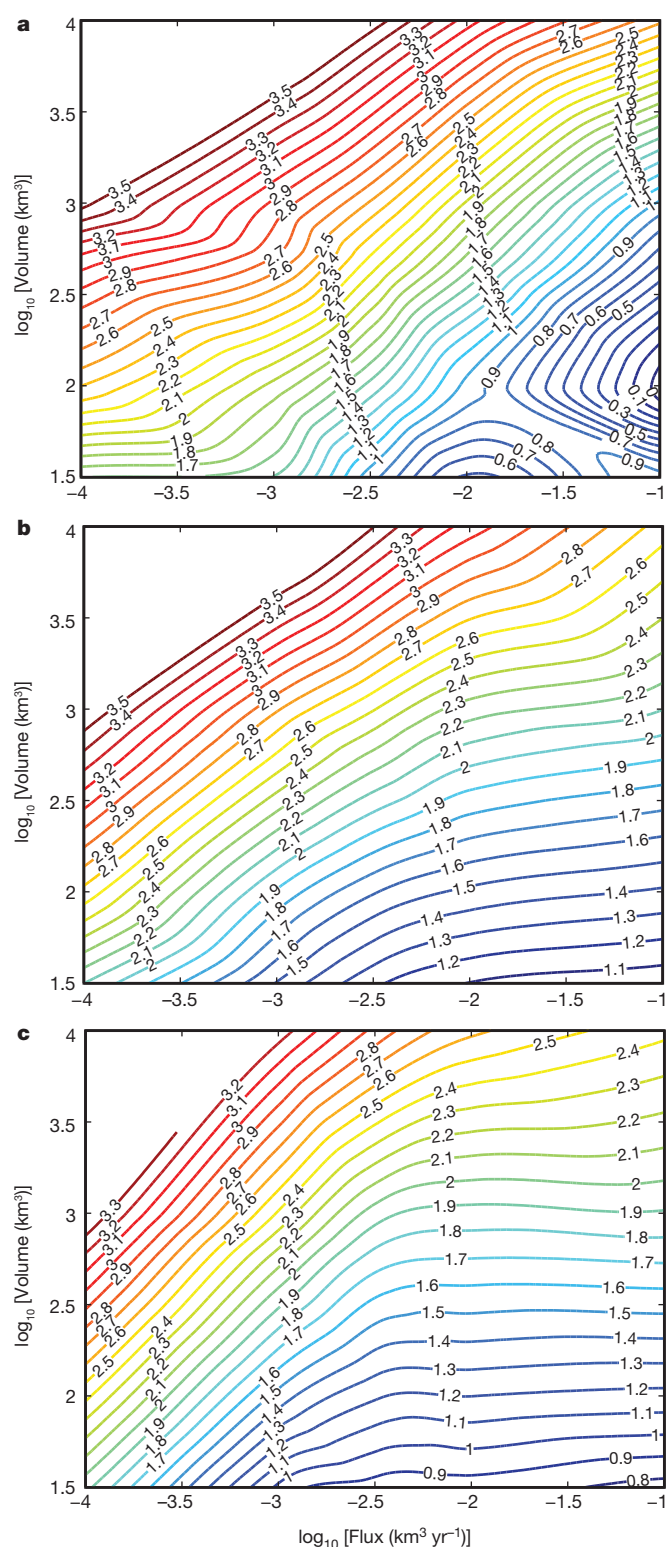
Figure 3 | Contours of mode, median and standard deviation for zircon crystallization time spectra, calculated from the numerical modelling. The contours are labelled with the logarithmic values of mode (**a**), median (**b**) and standard deviation (**c**) in kiloyears (kyr).

injected magma and the total ore endowment (larger for Bingham Canyon than for Coroccohuayco) shows the importance of determining the total volume of injected magma associated with mineralization in cases where the plutonic part of the system is not fully exposed.

The zircons of Coroccohuayco were collected in porphyry plugs that represents a volumetrically trivial portion of the larger magmatic system
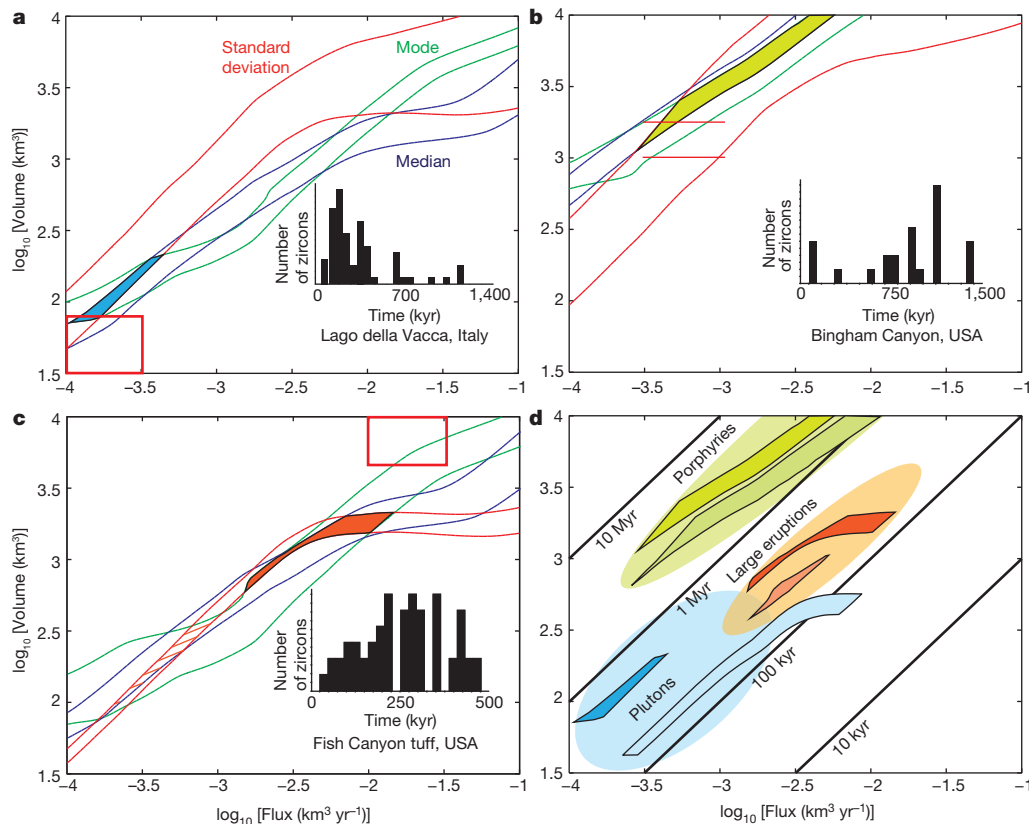
Figure 4 | Results of the inversion of natural populations of zircon ages.
a–c, The intersections of mode, median and standard deviation ranges provide best estimates of magma fluxes and volumes (shaded areas). Red boxes show volume and fluxes estimated using independent methods[5,6,8,9,15,22,23,31]. No flux estimates exist for Bingham Canyon (no red vertical lines). The inclined orange lines indicate the second region of overlap (excluded, see main text). Insets show the populations of zircon crystallization times for each system analysed (50-kyr bins and 54 analyses for **a**, 100-kyr bins and 22 analyses for **b**, and 25-kyr bins and 24 analyses for **c**). **d,** The coloured areas are the same as in **a**, **b** and **c** plus three additional systems (a porphyry, a large eruption and a pluton; Extended Data Fig. 3). The shaded areas highlight volume–flux regions specific for porphyries, large eruptions and plutons.

present at depth[25], yet our analysis of zircon ages provides an estimate of the total pluton volume and average magma flux that is comparable with those given by ref. 25 (Extended Data Fig. 3b). Zircons retrieved in volcanic products should also provide estimates of the volume of magma from which they crystallized before eruption and the average magma flux at which the magmatic reservoir was assembled.

To test this hypothesis, we considered a data set for the Fish Canyon tuff[26], which is the largest described volcanic eruption on the Earth[5,27]. Our inversion method provides two potential regions of magma flux and volume (Fig. 4b). We exclude the lower estimate because this provides a magma volume significantly smaller than the erupted volume and because the higher estimate is closer to the estimated erupted volume[27] (5,000 km³; Fig. 4b). Our method also provides good results for the Oruanui eruption[28] (New Zealand; Extended Data Fig. 3c). These data support the suggestion that large volumes of potentially eruptible magma accumulate in the crust at high average rates[3,28]. The similar fluxes obtained for the Oruanui eruption and the upper estimates for the Torres del Paine intrusion (Fig. 4d) also clearly show that the average magma flux is only one of the factors influencing the probability that magma will erupt at the surface[3,29,30].

Our results permit the quantitative determination of the relationship between magma fluxes and the long-term evolution of magmatic systems, which until now could only be qualitatively established. The currently available data suggest that: (1) short periods of magma injections in the crust at low injection rates leads to the construction of plutons[3,12,14,24]; (2) prolonged magmatic activity at relatively low magma injection rates seems to be a favourable condition for the genesis of ore deposits, and (3) relatively high magma fluxes over a restricted time are essential to construct the magmatic reservoirs feeding large volcanic eruptions[3] (Fig. 4d).

Our method requires a single set of analytical data, and provides opportunities to quantify magmatic fluxes at the global scale and to determine the evolution of magma productivity over the history of the Earth. Establishing these relationships will greatly improve our understanding of how global heat fluxes control the recurrence period of volcanic eruptions of different magnitudes, the accretion of the continental crust and the distribution of resources associated with magmatic activity on the Earth.

1. Crisp, J. A. Rates of magma emplacement and volcanic output. *J. Volcanol. Geotherm. Res.* **20,** 177–211 (1984).
2. White, S. M., Crisp, J. A. & Spera, F. J. Long-term volumetric eruption rates and magma budgets. *Geochem. Geophys. Geosyst.* **7,** Q03010 (2006).
3. Caricchi, L., Annen, C., Blundy, J., Simpson, G. & Pinel, V. Frequency and magnitude of volcanic eruptions controlled by magma injection and buoyancy. *Nature Geosci.* **7,** 126–130 (2014).
4. Longo, A. A., Dilles, J. H., Grunder, A. L. & Duncan, R. Evolution of calc-alkaline volcanism and associated hydrothermal gold deposits at Yanacocha, Peru. *Econ. Geol.* **105,** 1191–1241 (2010).
5. Lipman, P. W. Incremental assembly and prolonged consolidation of Cordilleran magma chambers: evidence from the Southern Rocky Mountain volcanic field. *Geosphere* **3,** 42 (2007).
6. Watson, E. B. & Harrison, T. M. Zircon saturation revisited: temperature and composition effects in a variety of crustal magma types. *Earth Planet. Sci. Lett.* **64,** 295–304 (1983).
7. Schaltegger, U. *et al.* Zircon and titanite recording 1.5 million years of magma accretion, crystallization and initial cooling in a composite pluton (southern Adamello batholith, northern Italy). *Earth Planet. Sci. Lett.* **286,** 208–218 (2009).
8. Schoene, B. *et al.* Rates of magma differentiation and emplacement in a ballooning pluton recorded by U–Pb TIMS-TEA, Adamello batholith, Italy. *Earth Planet. Sci. Lett.* **355–356,** 162–173 (2012).

9.   Michel, J., Baumgartner, L., Putlitz, B., Schaltegger, U. & Ovtcharova, M. Incremental growth of the Patagonian Torres del Paine laccolith over 90 k.y. *Geology* **36,** 459 (2008).
10. Leuthold, J. *et al.* Time resolved construction of a bimodal laccolith (Torres del Paine, Patagonia). *Earth Planet. Sci. Lett.* **325–326,** 85–92 (2012).
11. Lissenberg, C. J., Rioux, M., Shimizu, N., Bowring, S. A. & Mevel, C. Zircon dating of oceanic crustal accretion. *Science* **323,** 1048–1050 (2009).
12. Annen, C. From plutons to magma chambers: thermal constraints on the accumulation of eruptible silicic magma in the upper crust. *Earth Planet. Sci. Lett.* **284,** 409–416 (2009).
13. Michaut, C. & Jaupart, C. Ultra-rapid formation of large volumes of evolved magma. *Earth Planet. Sci. Lett.* **250,** 38–52 (2006).
14. Glazner, A. F., Bartley, J. M., Coleman, D. S., Gray, W. & Taylor, R. Z. Are plutons assembled over millions of years by amalgamation from small magma chambers? *GSA Today* **14,** 4–11 (2004).
15. Caricchi, L., Annen, C., Rust, A. & Blundy, J. Insights into the mechanisms and timescales of pluton assembly from deformation patterns of mafic enclaves. *J. Geophys. Res.* **117,** B11206 (2012).
16. Piwinskii, A. J. & Wyllie, P. J. Experimental studies of igneous rock series—a zoned pluton in Wallowa batholith Oregon. *J. Geol.* **76,** 205–234 (1968).
17. Boehnke, P., Watson, E. B., Trail, D., Harrison, T. M. & Schmitt, A. K. Zircon saturation re-revisited. *Chem. Geol.* **351,** 324–334 (2013).
18. Harrison, T. M., Watson, E. B. & Aikman, A. B. Temperature spectra of zircon crystallization in plutonic rocks. *Geology* **35,** 635 (2007).
19. Jaeger, J. C. Thermal effects of intrusions. *Rev. Geophys. Space Phys.* **2,** 443–466 (1964).
20. Marsh, B. D. On the crystallinity, probability of occurrence, and rheology of lava and magma. *Contrib. Mineral. Petrol.* **78,** 85–98 (1981).
21. von Quadt, A. *et al.* Zircon crystallization and the lifetimes of ore-forming magmatic-hydrothermal systems. *Geology* **39,** 731–734 (2011).
22. Steinberger, I., Hinks, D., Driesner, T. & Heinrich, C. A. Source plutons driving porphyry copper ore formation: combining geomagnetic data, thermal constraints, and chemical mass balance to quantify the magma chamber beneath the Bingham Canyon deposit. *Econ. Geol.* **108,** 605–624 (2013).
23. John, B. E. & Blundy, J. D. Emplacement-related deformation of granitoid magmas, southern Adamello massif, Italy. *Geol. Soc. Am. Bull.* **105,** 1517–1541 (1993).
24. de Saint Blanquat, M. *et al.* Multiscale magmatic cyclicity, duration of pluton construction, and the paradoxical relationship between tectonism and plutonism in continental arcs. *Tectonophysics* **500,** 20–33 (2011).
25. Chelle-Michou, C., Chiaradia, M., Ovtcharova, M., Ulianov, A. & Wotzlaw, J.-F. Zircon petrochronology reveals the temporal link between porphyry systems and the magmatic evolution of their hidden plutonic roots (the Eocene Coroccohuayco deposit, Peru). *Lithos* **198–199,** 129–140 (2014).
26. Wotzlaw, J. F. *et al.* Tracking the evolution of large-volume silicic magma reservoirs from assembly to supereruption. *Geology* **41,** 867–870 (2013).
27. Lipman, P. W., Dungan, M. A., Brown, L. L. & Deino, A. Recurrent eruption and subsidence at the Platoro caldera complex, southeastern San Juan volcanic field, Colorado: new tales from old tuffs. *Geol. Soc. Am. Bull.* **108,** 1039–1055 (1996).
28. Wilson, C. J. N. & Charlier, B. L. A. Rapid rates of magma generation at contemporaneous magma systems, Taupo volcano, New Zealand: insights from U-Th model-age spectra in zircons. *J. Petrol.* **50,** 875–907 (2009).
29. Jellinek, A. M. & DePaolo, D. J. A model for the origin of large silicic magma chambers: precursors of caldera-forming eruptions. *Bull. Volcanol.* **65,** 363–381 (2003).
30. Gregg, P. M., de Silva, S. L., Grosfils, E. B. & Parmigiani, J. P. Catastrophic caldera-forming eruptions: thermomechanics and implications for eruption triggering and maximum caldera dimensions on Earth. *J. Volcanol. Geotherm. Res.* **241–242,** 1–12 (2012).
31. Hanchar, J. M. & Watson, E. B. Zircon saturation thermometry. *Rev. Mineral. Geochem.* **53,** 89–112 (2003).

Author Contributions L.C. structured the study, took the lead on writing the manuscript, performed the statistical analysis of the data, and collected literature data. G.S. performed the numerical modelling and analysed the results. U.S. focused on the zircon geochronology. All authors jointly contributed to the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.C. (luca.caricchi@unige.ch).

METHODS

We performed thermal modelling to calculate the evolution of temperature in a crustal magmatic intrusion growing by continuous magma injection. The thermal modelling results were used to compute the number of zircons crystallizing over the history of the intrusion. This was achieved by considering a range of zircon crystallization temperatures of 700 °C–750 °C (T_{zr}) and computing the total number of new zircons crystallizing at each time step, assuming a constant rate of crystallization for each unit of magma within the range of T_{zr}. The sum of the number of zircons that crystallized at each time step within the magma body provided the total population of zircon crystallization times (Fig. 1).

This study is based on the observation that populations of zircon ages from different magmatic systems often show important variations in their statistical characteristics (Extended Data Fig. 1). We present the details of the thermal modelling we performed to establish the relationships between magma flux and final volume of injected magma and the mode, median and standard deviation of zircon age populations. We used models to test the sensitivity of our inversion method to parameters such as the range of crystallization temperature or geometry of magma injection (Extended Data Fig. 2). Finally, we provide additional results obtained from the inversion of zircon age spectra using our model (Extended Data Fig. 3).

Thermal modelling. Our results are based on numerical solutions of the two-dimensional axisymmetric diffusion equation for temperature obtained with the Petrov–Galerkin Finite Element Method on 4-node quadrilaterals. Our calculations include heat advection due to outward displacement of the wall rocks in response to magma emplacement and latent heat generation caused by crystallization in the magma chamber. Unless stated otherwise, all calculations were performed with the following physical parameters: initial geothermal gradient 30 °C km^{-1}, intrusion depth, 10 km, heat conductivity 2.7 W m^{-1} K^{-1}, heat capacity 1 kJ kg^{-1} K^{-1}, rock density 2,700 kg m^{-3} and latent heat of fusion 350 kJ kg^{-1}.

The crystal fraction (X_f) is assumed to vary as a function of temperature as:

$$X_f = 1 - \left(\frac{1}{1 + e^{\theta}} \right) \qquad (1)$$

where $\theta = \dfrac{(800 - T)}{23}$ and temperature T has units of °C (Extended Data Fig. 4). This equation was selected because it provides the best fit to the data of ref. 16 collected for granodioritic compositions. Most calculations were performed assuming outward-inflating spherical magma chambers, though we tested sensitivity to the intrusion geometry by comparing these results with horizontally expanding cylinders and vertically accreting cylinders (Supplementary Table 1). Simulations for volumes larger than 1,000 km^3 were all performed by vertically accreting the intrusion. The centre of the intrusion in these simulations was also fixed at a depth of 10 km. This geometry (instead of spherical) was necessary to avoid the intrusion of magma at extremely shallow depths or out of the crust. We studied a range of volumetric intrusive fluxes (0.0001–0.1 km^3 yr^{-1}) and emplacement durations between 0.3 kyr and 10,000 kyr.

Zircon crystallization. We considered the intrusion of granodioritic magma with an average zirconium concentration of 50–250 p.p.m. (http://georoc.mpch-mainz.gwdg.de). Given this bulk rock composition the range of zircon saturation temperature can be computed following the models of refs 6 and 17 to a range between 700 °C and 750 °C. This is a minimum range: considering the evolution of residual melt associated with crystallization could expand this range to higher temperatures[18]. To compute the number of zircons crystallized within the intrusion over time, we back-calculated the entire thermal history of each volume of magma and computed the histogram for each unit volume, by assuming that one zircon crystallizes every kiloyear (one time step) for the time this unit of magma volume spent at zircon saturation. We obtained the total population of zircon crystallization times by summing the populations of zircons obtained for each unit volume of magma. This final operation tends to generate high-frequency spikes in the calculated populations of zircon crystallization times observed in Figs 1 and 2. Note that considering a different number of zircons crystallizing in a unit volume of magma would change the total number of zircons but not the mode, median and standard deviation of the zircon populations. The resulting mode, median and standard deviation of the calculated zircon populations vary consistently as a function of magma flux and total volume of the intrusion as shown in Figs 2 and 3 and Supplementary Table 1.

Inversion method and model sensitivity to assumptions. From our model results, we have computed the mode, median and standard deviation of numerous zircon populations as a function of magma flux and final volume of the magma body (Fig. 3). Using these results, we can then do the inverse: given a measured mode, median and standard deviation for a natural (or model) zircon age population, we can back-compute the magma flux and final volume of injected magma, since in theory the three curves corresponding to these parameters intersect at one point in flux–volume space (Fig. 3). We have tested the validity of our approach by applying our inversion method to the populations of zircon ages calculated by the thermal model. Each inversion gave unique (that is, one intersection point between mode, median and standard deviation of the population) values of flux and volume that were identical to those used for the thermal modelling (Supplementary Table 1). The application of our inversion method to natural populations of zircon ages does not provide a single solution but rather a region of overlap between the minimum and maximum values (calculated with the bootstrap method at 95% confidence level) of mode, median and standard deviation of the population (Fig. 4 and Extended Data Fig. 3). The range in mode, median and standard deviation are a direct consequence of analytical uncertainties and the number of analyses. The range of mode, median and standard deviation are shown in Fig. 4a, b and c by the regions between the green, blue and red lines, respectively. To determine the effect of a larger range of zircon crystallization temperature on the final results of our model we performed calculations considering zircon saturation between 700 °C and 800 °C (the runs marked '2*' are performed considering a temperature interval for zircon crystallization between 700 °C and 800 °C). Comparing run numbers R41 with R41_2* and R45 with R45_2* highlight the small effect of magma composition, expressed here as a larger interval of zircon saturation temperature, on the main characteristics of the calculated zircon populations (Supplementary Table 1). We apply our inversion method (based on calculations performed with 700 °C < T_{zr} < 750 °C; see Fig. 3) to the populations of zircon ages calculated with 700 °C < T_{zr} < 800 °C (the same procedure was followed for all the sensitivity tests). The inversion did not result in a single value for magma flux and volume but rather a region of overlap with minimum and maximum values of mode, median and standard deviation for the calculated population of zircon ages (values in parentheses in Supplementary Table 1). However, the estimated magma flux and final volume of injected magma are within ±0.25 log units of the input values (Supplementary Table 1). Therefore, our method is potentially applicable to magmas with larger ranges of zircon crystallization temperature (that is, different chemistry and Zr concentration; ref. 18) than the 700 °C–750 °C interval considered here. We further tested the effect of higher temperature for the rocks surrounding the magma intrusion (for example, geothermal gradient of 50 °C km^{-1} instead of 30 °C km^{-1}) and also in this case the values obtained from the inversion are close to the input ones (R45_3*; Supplementary Table 1). Decreasing the temperature of injection of the magma leads to an overestimation of the recalculated magma flux and intrusion volume by 0.1 (that is, about 0.003 km^3 yr^{-1} on a flux of 0.01 km^3 yr^{-1}) and 0.15 log units (that is, about 40 km^3 of a total of 100 km^3), respectively.

To assess the effect of the modality of magma emplacement in the crust on the resulting populations of zircon crystallization ages, additional models were performed considering different geometries of magma input into the crust. Runs R60^{ch} and R61^{cv} were performed with the same input parameters of R42 but in R60^{ch} magma was injected at the core of the intrusion in cylinders and the magma body grew outwards horizontally, while in R61^{cv} the magma was injected in cylinders at the core of the intrusion that grew vertically (expansion towards shallower and deeper levels). Extended Data Figure 2 shows the similar temperature fields for these three runs after 100 kyr. This, together with the results in Supplementary Table 1, show that different intrusion geometries have a limited influence on the calculated populations of zircons. The largest difference (0.4 logarithmic units) is found between the mode calculated in R42 and R60^{ch}. This difference is similar to the range obtained from the statistical analysis of natural populations of zircon ages (Figs 3 and 4).

In our calculations we did not consider resorption of zircons during the assembly of the magmatic bodies in the crust. This last assumption implies that our model is valid for all magmatic bodies in which the injection of magma produces local temperature increases that are insufficient to cause resorption of a large proportion of the crystallized zircons. This is appropriate for magmatic bodies for which the mass of injected magma is smaller, or significantly smaller than the mass of resident magma. Studies of intrusive upper crustal magmatic bodies, where single pulses of magma are significantly smaller than the host volume[14,15,24], confirm the applicability of our model to most upper-crustal magmatic bodies. This implies that even if each temperature oscillation should lead to resorption of zircons in a limited portion of the intrusion, at the same time a much larger volume of magma would still continue to crystallize zircons. An additional consideration comes from the observation that zircons are often zoned. One isotope dilution thermal ionization mass spectrometry (ID-TIMS) bulk zircon analysis provides an average age that is close to the time at which most of the zircon mass crystallized. Because our approach is based on the statistical characterization of zircon populations, this implies that the modal age of a population of zircon ages still reflects the time at

which most of the zircons crystallized (or the largest mass of different zircons grew).

The assumptions made in our study target the construction of a model with a wide range of applicability to magmas crystallizing zircons in the upper continental crust. The same procedure can be applied to other sets of conditions designed to identify magma fluxes in other tectonic settings or depths in the crust. Our tests suggest that the effect of parameters such as T_{zr}, geothermal gradient or geometry of emplacement are generally smaller than the range of estimates of magma flux and volume associated to analytical uncertainty. Nevertheless, we suggest limiting the application of our method to granodioritic–tonalitic magmas that crystallized the highest possible number of zircons in the upper crust.

32. Piwinskii, A. J. & Wyllie, P. J. Experimental studies of igneous rock series. felsic body suite from Needle Point pluton, Wallowa-Batholith, Oregon. *J. Geol.* **78,** 52–76 (1970).
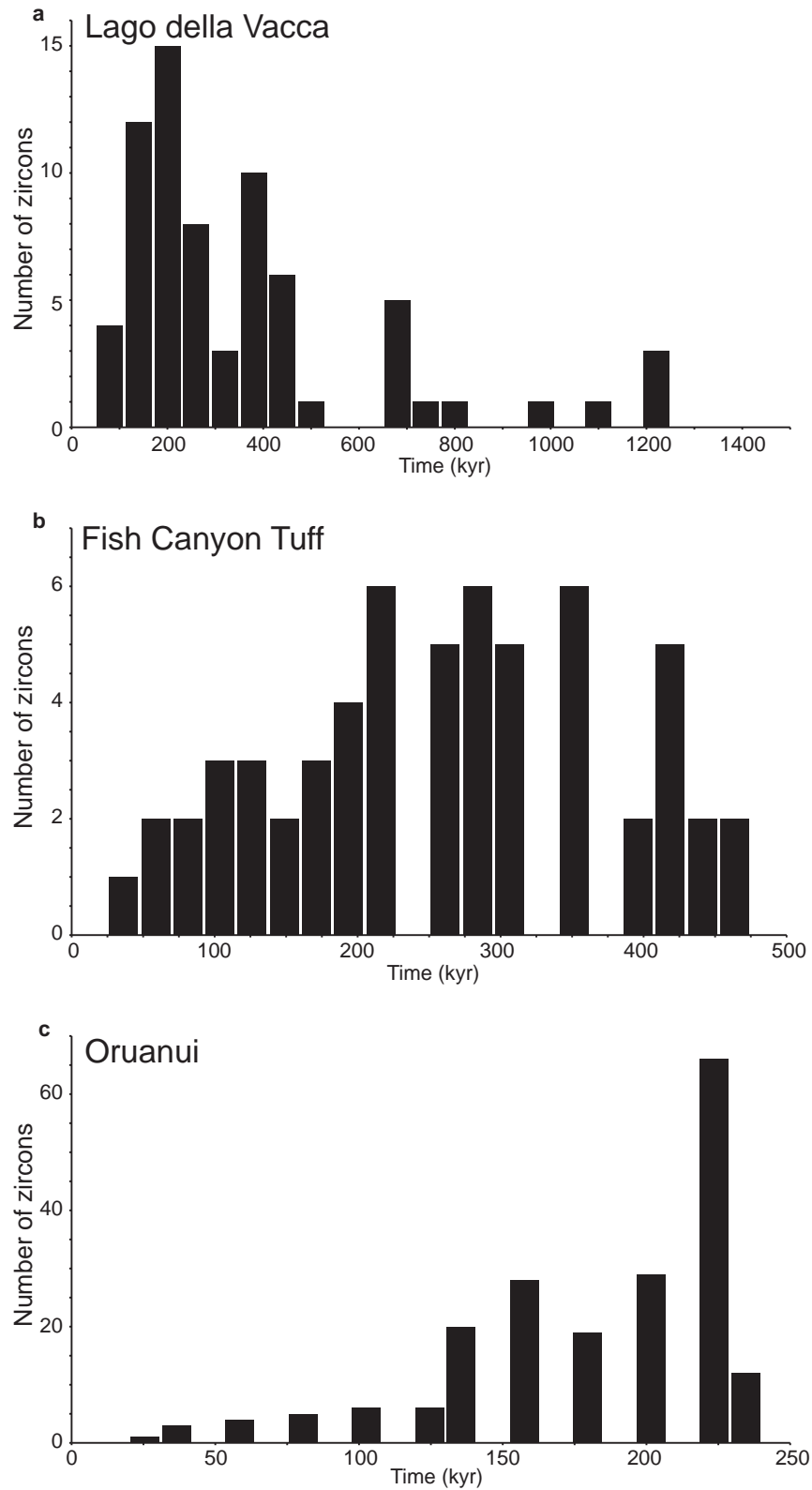
Extended Data Figure 1 | Distributions of zircon crystallization times.
a, The Lago della Vacca pluton[7,8] (in Italy); **b**, Fish Canyon Tuff eruption[26] (in USA); **c**, Oruanui eruptions[28] (in New Zealand). The zircon crystallization times are calculated by subtracting each zircon age from the age of the oldest zircon of the population.
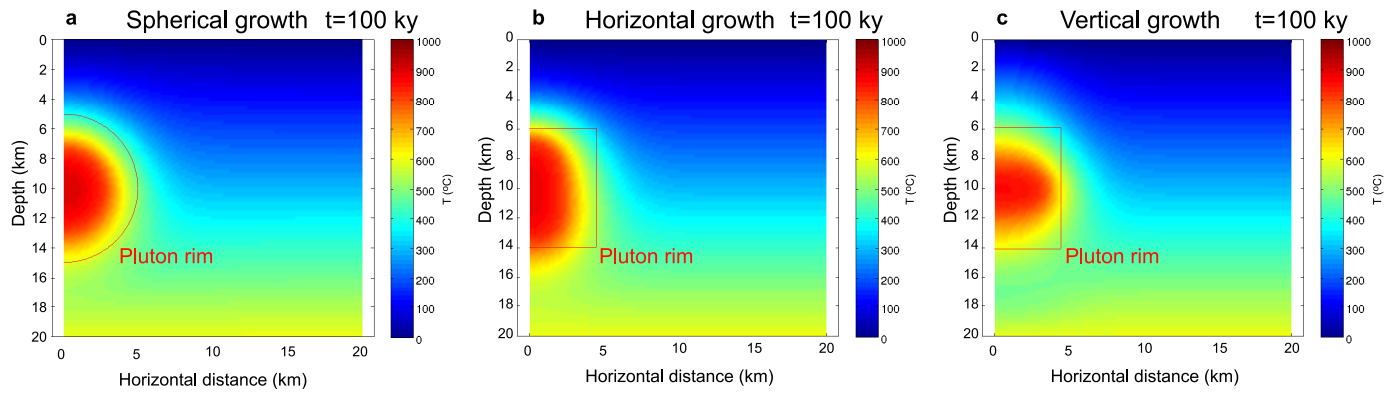
Extended Data Figure 2 | Distribution of temperature after 100 kyr of magma injection in magma bodies emplaced with different modalities. For all panels the rate of magma injection is $10^{-2}\,\mathrm{km^3\,yr^{-1}}$, the final volume of injected magma is $500\,\mathrm{km^3}$ and the initial wall rock temperature at 10 km is $300\,^{\circ}\mathrm{C}$. **a**, Magma is injected at the core. **b**, Magma is injected in vertically elongated pulses and the magma body grows by lateral displacement of the surrounding crust. **c**, Sill-like magma batches are stacked vertically and the intrusion grows by displacement of the surrounding crust along the vertical direction.
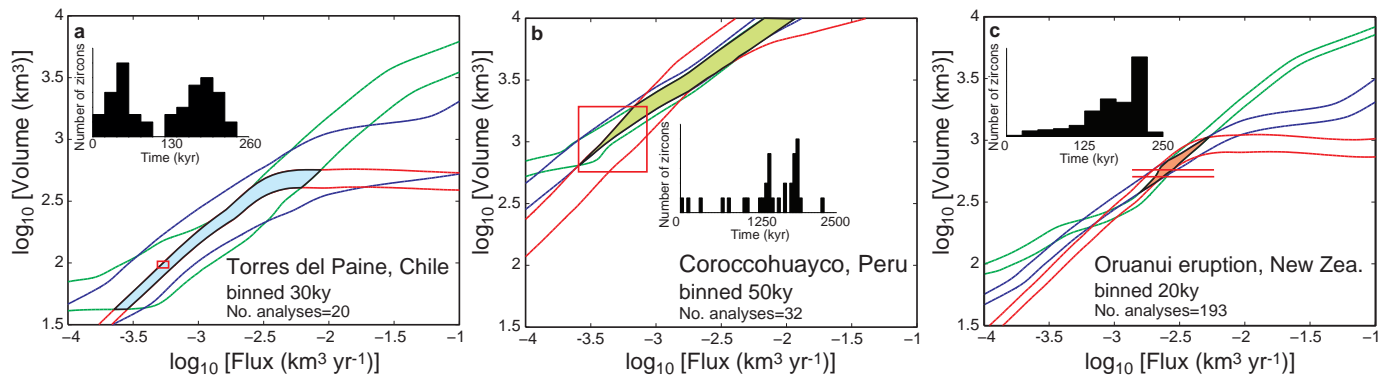
Extended Data Figure 3 | Results obtained by the inversion of zircon populations. **a**, Torres del Paine granites; **b**, The Coroccohuayco porphyry; **c**, Oruanui eruption. The range of estimated volume and magma fluxes are highlighted by the shaded areas and plotted in Fig. 4 for comparison with similar magmatic and volcanic systems. The red lines provide independent estimates of magma flux and final volume of the magmatic bodies. Estimates for the average magma flux into the system do not exist for the Oruanui eruption and therefore the vertical lines of the red box were not traced. Data are from refs 9 and 10 for Torres del Paine, ref. 25 for Coroccohuayco and ref. 28 for Oruanui. The insets in the figure show the populations of zircon crystallization times on which the statistical analysis has been performed.
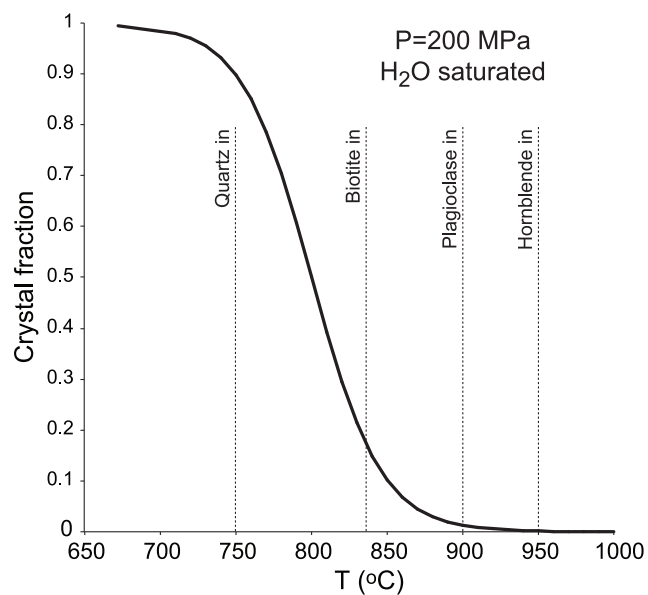
Extended Data Figure 4 | Variation of crystal fraction as a function of temperature. The curve is calculated using equation (1), which provides the best fit to the data of ref. 32 collected at a confining pressure of 200 MPa and water-saturated conditions. The temperatures at which different phases appear in the crystallizing assemblage are from ref. 16.

LETTER

Climate change selects for heterozygosity in a declining fur seal population

Jaume Forcada[1] & Joseph Ivan Hoffman[2]

Global environmental change is expected to alter selection pressures in many biological systems[1–3], but the long-term molecular and life history data required to quantify changes in selection are rare[4]. An unusual opportunity is afforded by three decades of individual-based data collected from a declining population of Antarctic fur seals in the South Atlantic. Here, climate change has reduced prey availability and caused a significant decline in seal birth weight. However, the mean age and size of females recruiting into the breeding population are increasing. We show that such females have significantly higher heterozygosity (a measure of within-individual genetic variation) than their non-recruiting siblings and their own mothers. Thus, breeding female heterozygosity has increased by 8.5% per generation over the last two decades. Nonetheless, as heterozygosity is not inherited from mothers to daughters, substantial heterozygote advantage is not transmitted from one generation to the next and the decreasing viability of homozygous individuals causes the population to decline. Our results provide compelling evidence that selection due to climate change is intensifying, with far-reaching consequences for demography as well as phenotypic and genetic variation.

Heterozygosity is often associated with Darwinian fitness in natural populations[5]. Episodic heterozygote advantage can lead to an increase in the proportion of heterozygous genotypes during periods of environmental stress, particularly in populations undergoing demographic changes[6]. However, heterozygosity usually has low heritability (although dissenting results have been reported[7]) and it is unclear whether short-term changes in the proportion of heterozygous individuals can generate measurable impacts on life history and population parameters, which in turn could help to buffer the impacts of long-term climate change.

An unprecedented opportunity to investigate the relationship between individual heterozygosity and demographic response to climate change is afforded by three decades of biometric, life history and genetic data from a population of Antarctic fur seals (*Arctocephalus gazella*). In this species, several key components of fitness including male reproductive success, body size and mate choice correlate with heterozygosity measured at nine microsatellite markers[8–10]. We analysed 31 years of individual-based data (1982–2012 inclusive) from a population on the island of South Georgia, in an area of the southwest Atlantic strongly affected by climate change[11–14]. This population shows significant demographic fluctuation[12,14] linked to two major modes of climate variation: El Niño Southern Oscillation and the Southern Annular Mode (SAM). The SAM impacts the local ecosystem more directly through atmospheric, cryospheric and oceanographic variability[13], in turn modulating the availability of Antarctic krill (*Euphausia superba*), a keystone species and staple food of these fur seals[11,14].

Antarctic fur seals became abundant after recovering from near extinction in the nineteenth century and are major consumers of krill in the southwest Atlantic. However, the SAM has been mostly positive and increasing since the early 1990s, and positive SAM conditions are typically associated with high sea surface temperature and reduced krill availability[14]. We detected a 24% (standard error of the mean (s.e.m.) = 8%) decline in the number of breeding females over the past 27 years and

annual female numbers are accurately predicted by a model based on SAM conditions and krill availability (Fig. 1a), suggesting a tight linkage between climate change, food availability and population size. The period 2003–12 encompasses the worst observed population crashes and has witnessed a decline in female numbers of around 30%.

To explore this species' response to climate change, we analysed biometric data and used population models based on estimated vital rates including survival, recruitment and fecundity. Female life histories were modelled as a series of breeding decisions conditional on survival and the SAM (a proxy for food availability). Models were parameterized using mark–recapture data and include a locus-weighted measure of homozygosity (HL)[15] based on nine microsatellite markers (Extended Data Fig. 1 and Extended Data Table 1). Simultaneous effects of the SAM and HL were investigated for the period 2000–12 using the individual breeding histories of 1,703 closely monitored females of which 760 were genotyped, corresponding to 11 breeding cohorts of which 6 were expected to be fully recruited (that is, all of the females were older than the maximum age of first breeding by 2012).

Over the last 27 years, the average birth weight of female pups declined by 7.8% (Fig. 1a). However, since 2003, females recruiting into the breeding population were 5% heavier at birth compared with non-recruited females (Fig. 1b). They were also on average almost a year older (Fig. 1c) and with a larger body size for their age when recruited compared with two decades earlier when the population was stable (Fig. 1d). These results support findings from other long-lived mammals, where sustained food stress reduces both juvenile survival and reproductive rates, with a concomitant rise in age of first reproduction[16].

Along with changes in size and age of first breeding, the mean HL of breeding females also declined significantly over the last two decades (Fig. 2a). This decline is estimated at 17% overall, or approximately 8.5% per mean generation time, which itself decreased from 13.7 to 6.1 years as adverse SAM conditions are also associated with reduced female longevity. This long-term trend indicates increasingly strong selection against relatively homozygous individuals over time, as females recruiting after 2001 are significantly less homozygous than their own mothers, whereas non-recruited females are not (Fig. 2b). These changes appear to be unrelated to mate choice[9] since the mean heterozygosity of female offspring does not change significantly over time (F-test for linear regression, $F_{1,978} = 0.518$, $P = 0.472$). Instead, it appears that, as food availability declines, increasingly only the most heterozygous females are able to achieve breeding condition.

To further understand how heterozygosity has increased over time, we examined SAM and HL effects on vital rates since 2000 (Supplementary Information). Variation in SAM significantly affects most of the life cycle (Fig. 2c, d and Extended Data Fig. 2). During extreme positive SAM anomalies, mean juvenile survival, adult survival and fecundity declined by up to 37.1% (s.e.m. = 7.1%), 41.5% (5.1%) and 32.1% (8.5%) respectively. Survival of pre-breeders was inversely related to HL (Extended Data Fig. 2a), particularly in first-year females. Even stronger associations are seen between HL and the probability of recruiting at a given age, with relatively homozygous females often failing to recruit, especially under

[1]British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK. [2]Department of Animal Behaviour, University of Bielefeld, Postfach 100131, 33501 Bielefeld, Germany.
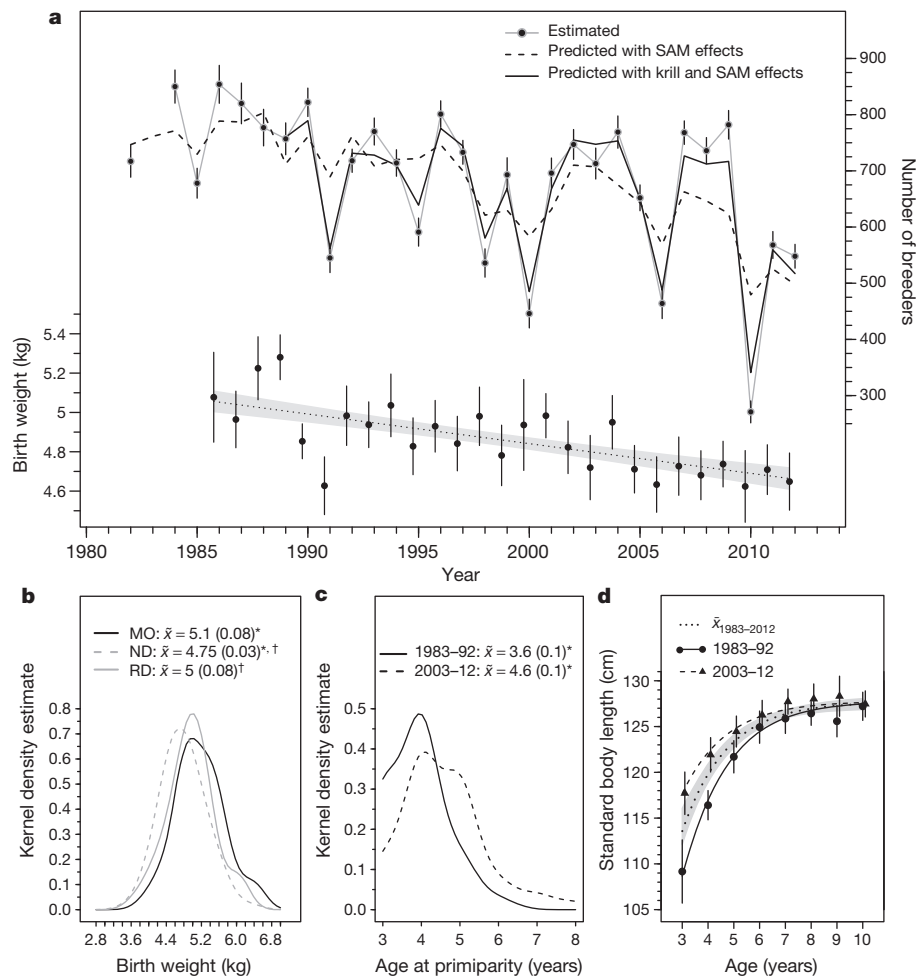
Figure 1 | Long-term population and phenotypic change associated with the Southern Annular Mode (SAM) in females of an Antarctic fur seal population. a, Population growth rate model, best predicted by the SAM and Antarctic krill availability (solid black line). **a–d**, Directional phenotypic response in birth weight (**a**, **b**; $n = 1606$ pups), age at primiparity (**c**; $n_{1983–92} = 102$ and $n_{2003–12} = 100$) and body size at age (**d**; $n_{1983–92} = 238$ and $n_{2003–12} = 256$). In **b** and **c**, medians ($\bar{x}$) are given, and identical superscript symbols indicate significant differences ($P < 0.05$, Mood's median test; median

absolute deviations in parentheses). MO are mothers ($n = 54$), ND daughters never seen recruiting ($n = 386$) and RD recruited daughters ($n = 42$). 1983–92 and 2003–12 (**c**, **d**) were periods where consistent measurements allowed independent decadal comparisons for a stable and a rapidly declining population, respectively. Grey shaded areas and vertical bars correspond to 95% CIs. Solid and dashed lines in **d** are fitted Gompertz growth models including the predicted mean (dotted line) of the two periods.

worse SAM conditions (Fig. 2c, d). HL also strongly influences a mother's ability to rear a pup, depending on her age and breeding history. In particular, females with high HL that failed or deferred breeding previously were more likely to fail subsequent breeding attempts than their more heterozygous counterparts (Fig. 2e). Relatively homozygous mothers therefore contribute offspring with lower viability to the population.

For a population to persist, its growth rate (λ) must be one or larger[17]. In long-lived species, the vital rates that contribute most to variation in λ usually buffer populations against environmental variability[14]. We estimated the stochastic form of λ, expressed as $\log(\lambda_s)$[17], finding it to be most sensitive to changes in the survival of breeders and pre-breeders (Fig. 3 and Extended Data Table 2). However, $\log(\lambda_s)$ was also sensitive to variation in juvenile survival rates, which in turn formed the largest contributory term to temporal variation in growth rate ($\hat{\sigma}_\lambda$). As expected, the SAM greatly influenced these vital rates and contributed most to the observed change in growth rate ($\Delta\lambda$) during the period 2002–12 (Extended Data Table 2), where $\bar{\lambda}$ was 0.903 ($\hat{\sigma}_\lambda = 0.09$) and hence the population was in decline (Fig.1a). In contrast, the strongest HL effects were seen on recruitment and breeding success, which only marginally contributed to $\hat{\sigma}_\lambda$ and to which $\log(\lambda_s)$ was insensitive. In absolute terms, the SAM contributed 87.5% to $\Delta\lambda$ and HL 12.5%. This nevertheless represents an unprecedentedly high contribution of heterozygosity to population growth rate[18].

The population mean HL estimated using integral projection models[19,20] was most sensitive to changes in parameters in the inheritance function (Extended Data Fig. 4), which describes the probability distribution of HL in pups as a function of the HL of mothers. However, neither these parameters nor the heritability of HL, estimated as mother–offspring regression of HL, were significantly different from zero (Supplementary Information). In agreement, the strength of viability selection, defined as the percentage contribution of survival to change in HL, was significantly higher (1.14%; confidence interval (CI): 0.50; 1.75) than the strength of fertility selection, defined as the percentage contribution of fertility to change in HL (0.51%; CI: −0.13; 1.02), and was most sensitive to changes in SAM effects on the survival of pre-breeders (Extended Data Fig. 5). Moreover, mean estimates of viability selection on HL significantly increased with increased SAM (Fig. 3a) and over time (Fig. 3b) as change towards adverse SAM conditions altered the functional relationship between survival and heterozygosity. This supports the hypothesis that with worsening SAM conditions relatively homozygous individuals are progressively filtered out of the population early in life, primarily through viability selection.

Many previous studies have reported heterozygosity–fitness correlations, but rarely does r^2 exceed 5% and it is more often around 1% (ref. 5). There are several possibilities as to why we uncovered such a strong effect. First, by integrating high-precision, long-term, individual-based data with
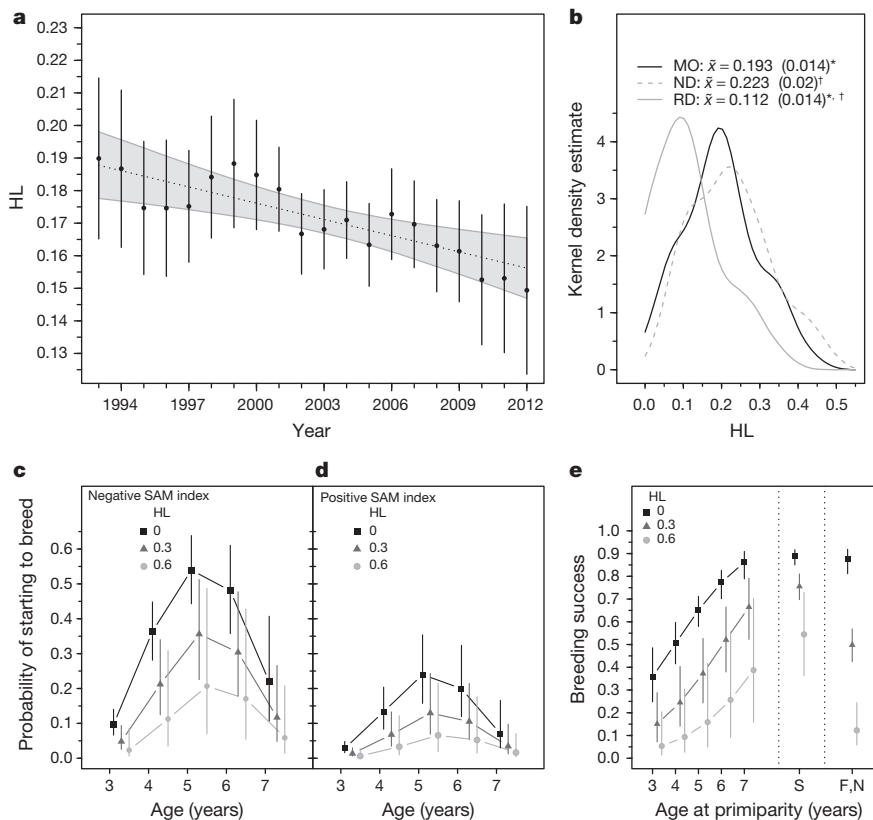
Figure 2 | Long-term trend in breeding female homozygosity, expressed as homozygosity weighted by locus (HL), and the effects of homozygosity on female recruitment and breeding success. **a**, Temporal trend in HL ($n = 713$ females) with linear regression fit (dotted line with 95% CI in shaded area). **b**, Differences in HL between mothers (MO, $n = 45$) and their recruited daughters (RD, $n = 40$) and non-recruited daughters (ND, $n = 38$); median estimates ($\tilde{x}$) are given and median absolute deviations are in parentheses; values with the same superscript symbol differ significantly ($P < 0.05$, Mood's median test). **c, d**, Variation in probability of starting breeding at age (recruitment) under negative and positive SAM values, evaluated at low (0), intermediate (0.3) and high (0.6) homozygosity (HL). **e**, Effects of variation in HL on breeding success ; S indicates females that successfully bred the previous year and F, N indicates females that previously failed or deferred breeding. All vertical bars show 95% CIs.
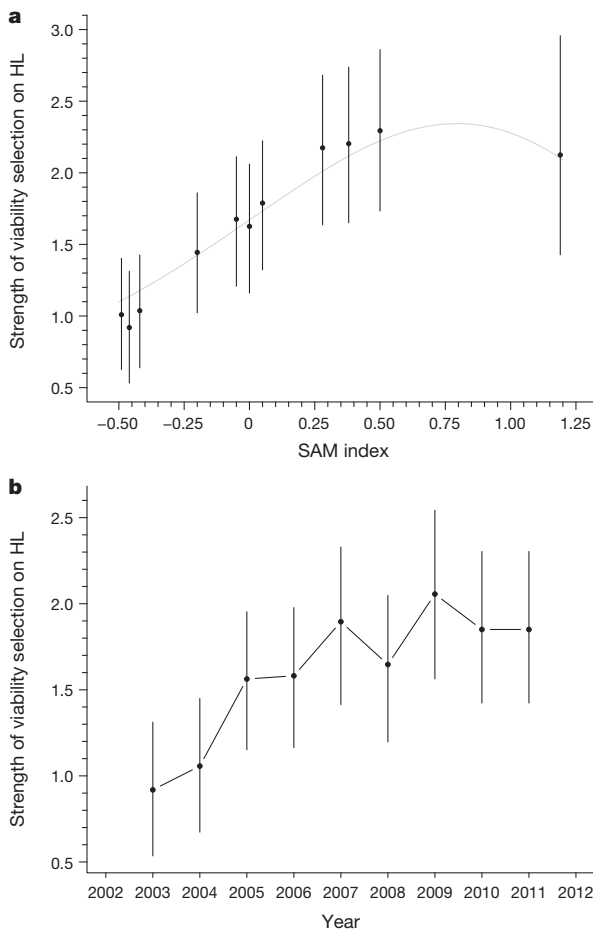
strongly relevant climate data in a well-defined analytical framework, we can achieve unusually accurate phenotyping. Second, environmental stress is known to exacerbate inbreeding depression[21] and our declining study population is clearly stressed. Third, the combination of polygyny and natal site fidelity in fur seals probably creates a much greater risk of inbreeding relative to many other species. Accordingly, six out of nine loci show individually significant long-term declines in HL (Extended Data Fig. 6) and the difference in HL between mothers and their recruited daughters is significant for four loci (Extended Data Table 3), consistent with a genome-wide reduction in heterozygosity rather than one or two loci driving the pattern. Moreover, for a subset of pups genotyped at an additional 101 genome-wide distributed single nucleotide polymorphisms[22] multilocus HL is significantly correlated between the two marker sets (F-test for linear regression, $F_{1, 347} = 80.78$, $P < 0.0001$), which is only to be expected if there is variance in inbreeding in the sample of individuals. The measure g_2, which quantifies the excess of double-heterozygotes at two loci relative to the expectation under random association, does not differ significantly from zero for the microsatellite data (0.001; s.d. = 0.002, $P = 0.19$; permutation test), but this probably reflects the fact that inbreeding is more readily detected through its effects on phenotype than through correlations in heterozygosity between individual pairs of loci[23]. The expected correlation between microsatellite heterozygosity and the inbreeding coefficient, calculated following the method described by ref. 23, is only -0.211, indicating that, if inbreeding could be estimated with greater precision, the effect size should be even larger[24].

Figure 3 | Strength of viability selection on homozygosity weighted by locus (HL) for the period 2002–12 inclusive. **a**, Variation with the SAM index, where points are estimated values and the fitted grey line indicates the prediction of the integral projection model. **b**, Running mean with a window of 3 years over the study period, which highlights the observed temporal increase in the mean accounting for year-to-year fluctuations; without a temporal increase, the mean should stay constant across years. Strength of selection is expressed as the percentage effect of survival on HL. All vertical bars show 95% CIs.

Predicting how species respond to climate change requires an understanding of the genetic basis of variation in individual fitness[18]. This is particularly important for long-lived species that may have relatively few generations in which to respond[2] and for species living in polar areas where rapid and profound changes are occurring[25]. As environmental conditions have deteriorated due to climate change, we explored how reduced food availability is exerting increasingly strong viability selection on an Antarctic fur seal population. Less fit, relatively homozygous females are increasingly excluded from the breeding population, reflected in a 17% increase in average breeding female heterozygosity over a 20-year period. Heterozygosity is not heritable from mothers to daughters, so the shift towards older, more heterozygous breeding females is not an evolutionary response to climate change. However, a more positive speculation is that, as environmental conditions continue to worsen, heterozygote advantage could promote the persistence of genetic variation[26], potentially buying time to allow the species to respond via adaptation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.
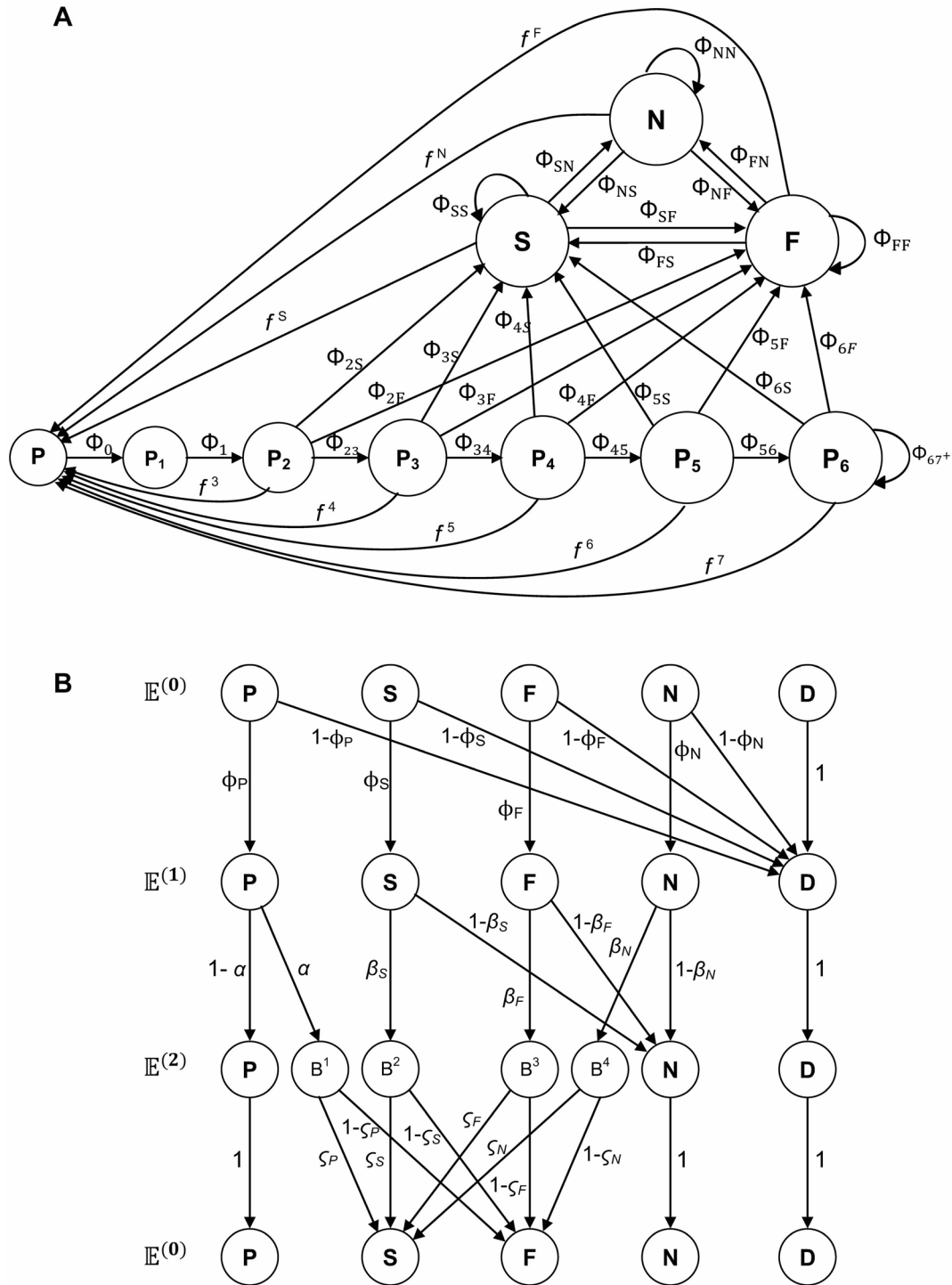
1. Parmesan, C. Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Evol. Syst.* **37,** 637–669 (2006).
2. Hoffmann, A. A. & Sgro, C. M. Climate change and evolutionary adaptation. *Nature* **470,** 479–485 (2011).
3. Franks, S. J. & Hoffman, A. A. Genetics of climate change adaptation. *Annu. Rev. Genet.* **46,** 185–208 (2012).
4. Gienapp, P., Teplitsky, C., Alho, J. S., Mills, J. A. & Merila, J. Climate change and evolution: disentangling environmental and genetic responses. *Mol. Ecol.* **17,** 167–178 (2008).
5. Chapman, J. R., Nakagawa, S., Coltman, D. W., Slate, J. & Sheldon, B. C. A quantitative review of heterozygosity-fitness correlations in animal populations. *Mol. Ecol.* **18,** 2746–2765 (2009).
6. Samollow, P. B. & Soulé, M. E. A case of stress related heterozygote superiority in nature. *Evolution* **37,** 646–649 (1983).
7. Mitton, J. B., Schuster, W. S. F., Cothran, E. G. & De Fries, J. C. The correlation between the individual heterozygosity of parents and their offspring. *Heredity* **71,** 59–63 (1993).
8. Hoffman, J. I., Boyd, I. L. & Amos, W. Exploring the relationship between parental relatedness and male reproductive success in the Antarctic fur seal *Arctocephalus gazella. Evolution* **58,** 2087–2099 (2004).
9. Hoffman, J. I., Forcada, J., Trathan, P. N. & Amos, W. Female fur seals show active choice for males that are heterozygous and unrelated. *Nature* **445,** 912–914 (2007).
10. Hoffman, J. I., Forcada, J. & Amos, W. Getting long in the tooth: a strong positive correlation between canine size and heterozygosity in the Antarctic fur seal *Arctocephalus gazella. J. Hered.* **101,** 527–538 (2010).
11. Murphy, E. J. *et al.* Climatically driven fluctuations in Southern Ocean ecosystems. *Proc. R. Soc. B* **274,** 3057–3067 (2007).
12. Forcada, J., Trathan, P. N. T., Reid, K. & Murphy, E. J. The effects of global climate variability in pup production of antarctic fur seals. *Ecology* **86,** 2408–2417 (2005).
13. Meredith, M. P., Murphy, E. J., Hawker, E. J., King, J. C. & Wallace, M. I. On the interannual variability of ocean temperatures around South Georgia, Southern Ocean: Forcing by El Niño/Southern Oscillation and the Southern Annular Mode. *Deep-Sea Res. II* **55,** 2007–2022 (2008).
14. Forcada, J., Trathan, P. N. & Murphy, E. J. Life history buffering in Antarctic mammals and birds against changing patterns of climate and environmental variation. *Glob. Change Biol.* **14,** 2473–2488 (2008).
15. Aparicio, J. M., Ortego, J. & Cordero, P. J. What should we weigh to estimate heterozygosity, alleles or loci? *Mol. Ecol.* **15,** 4659–4665 (2006).
16. Eberhardt, L. L. A paradigm for population analysis of long-lived vertebrates. *Ecology* **83,** 2841–2854 (2002).
17. Caswell, H. *Matrix Population Models* 2nd edn (Sinauer Associates, 2001).
18. Di Fonzo, M. M. I., Pelletier, F., Clutton-Brock, T. H., Pemberton, J. M. & Coulson, T. The population growth consequences of variation in individual heterozygosity. *PLoS ONE* **6,** e19667 (2011).
19. Easterling, M. R., Ellner, S. P. & Dixon, P. M. Size-specific sensitivity: applying a new structured population model. *Ecology* **81,** 694–708 (2000).
20. Coulson, T., Tuljapurkar, S. & Childs, D. Z. Using evolutionary demography to link life history theory, quantitative genetics and population ecology. *J. Anim. Ecol.* **79,** 1226–1240 (2010).
21. Jiménez, J. A., Hughes, K. A., Alaks, G., Graham, L. & Lacy, R. C. An experimental study of inbreeding depression in a natural habitat. *Science* **266,** 271–273 (1994).
22. Hoffman, J. I., Tucker, R., Clark, M. S., Forcada, J. & Slate, J. Rates of assay success and genotyping error when single nucleotide polymorphism genotyping in non-model organisms: a case study in the Antarctic fur seal. *Mol. Ecol. Resour.* **12,** 861–872 (2012).
23. Szulkin, M., Bierne, N. & David, P. Heterozygosity-fitness correlations: a time for reappraisal. *Evolution* **64,** 1202–1217 (2010).
24. Hoffman, J. I. *et al.* High-throughput sequencing reveals inbreeding depression in a natural population. *Proc. Natl Acad. Sci. USA* **111,** 3775–3780 (2014).
25. Meredith, M. P. & King, J. C. Rapid climate change in the ocean west of the Antarctic Peninsula during the second half of the 20th century. *Geophys. Res. Lett.* **32,** L19604 (2005).
26. Hoffmann, A. A. & Hercus, M. J. Environmental stress as an evolutionary force. *Bioscience* **50,** 217–226 (2000).

Extended Data Figure 1 | Antarctic fur seal life cycle and directed acyclic graph of multi-event mark–recapture models used to estimate vital rates. a, In the life cycle graph, nodes correspond to stages and arrows to probabilities of transition (ϕ_i) between stages (Supplementary Information), from year t to $t + 1$. Subscripts are for ages 0, for weanlings, to 6 at full physical maturity, and breeding stages. P_6 acts as a terminal node for individuals observed alive but never recruited. Fertilities (f) are female weanlings contributed by females breeding at $t + 1$. **b**, The acyclic graph shows intermediate sets of stages which are connected by rows, with each stage in a row being represented by a node. Different transitions are represented by arrows linking intermediate states. Nodes are different in **a** and **b**. See Supplementary Information for definitions of stages and vital rates.
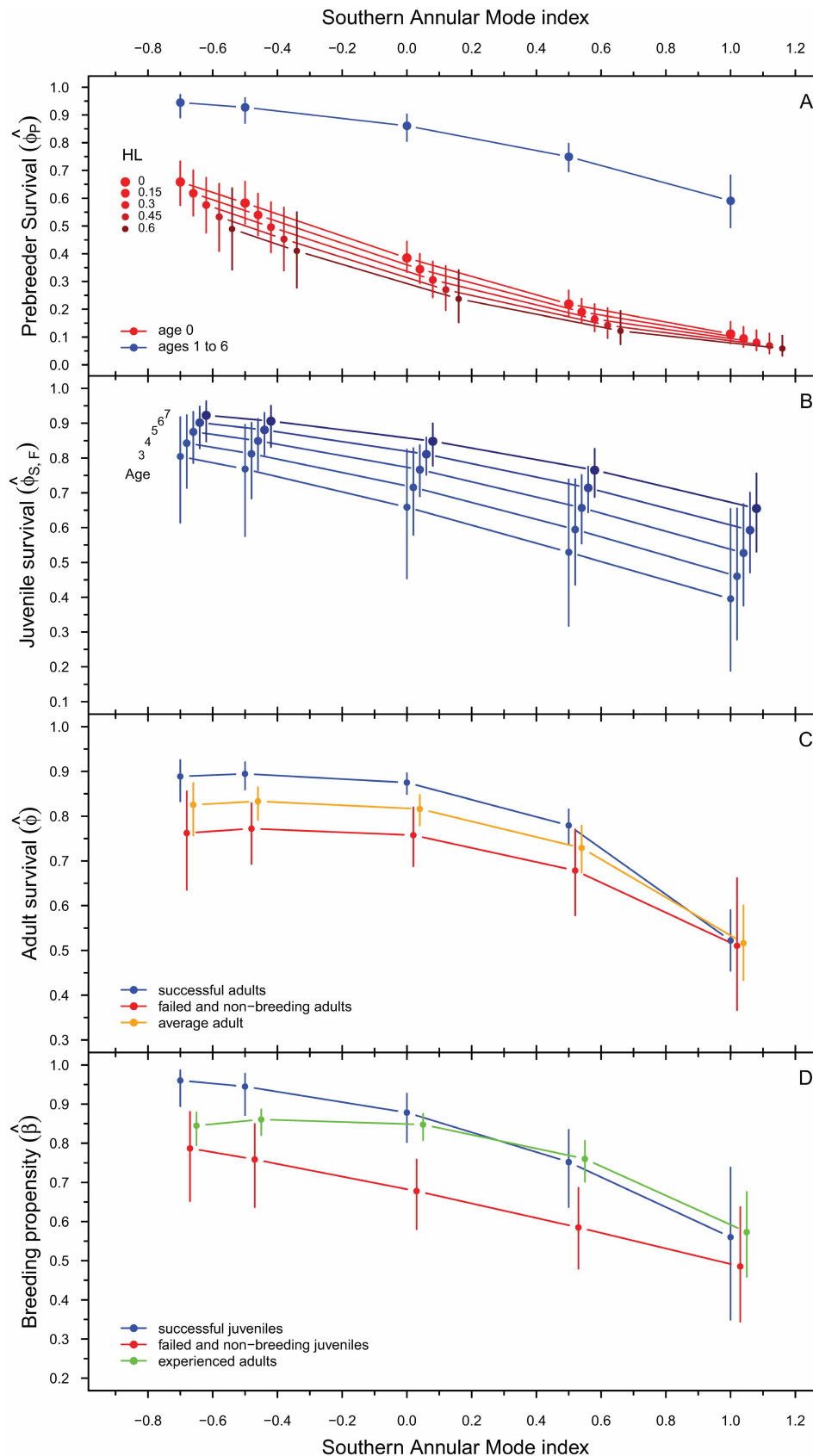
Extended Data Figure 2 | Variation in survival probability and breeding propensity (fecundity) with variation in the SAM index. Panel **a** shows survival of first-year pre-breeders in red, including its variation with homozygosity weighted by locus (HL); averaged survival for years 1 to 6 is shown in blue. Panel **b** shows age-specific survival for recruiting seals, and panel **c** the survival for adults (that is, seals with previous breeding experience) with variation by stages defined according to previous breeding outcomes (successful, failed, or deferred breeding). Panel **d** shows inter-annual breeding propensity for recently recruited seals and for experienced breeders. All vertical bars show 95% CIs.
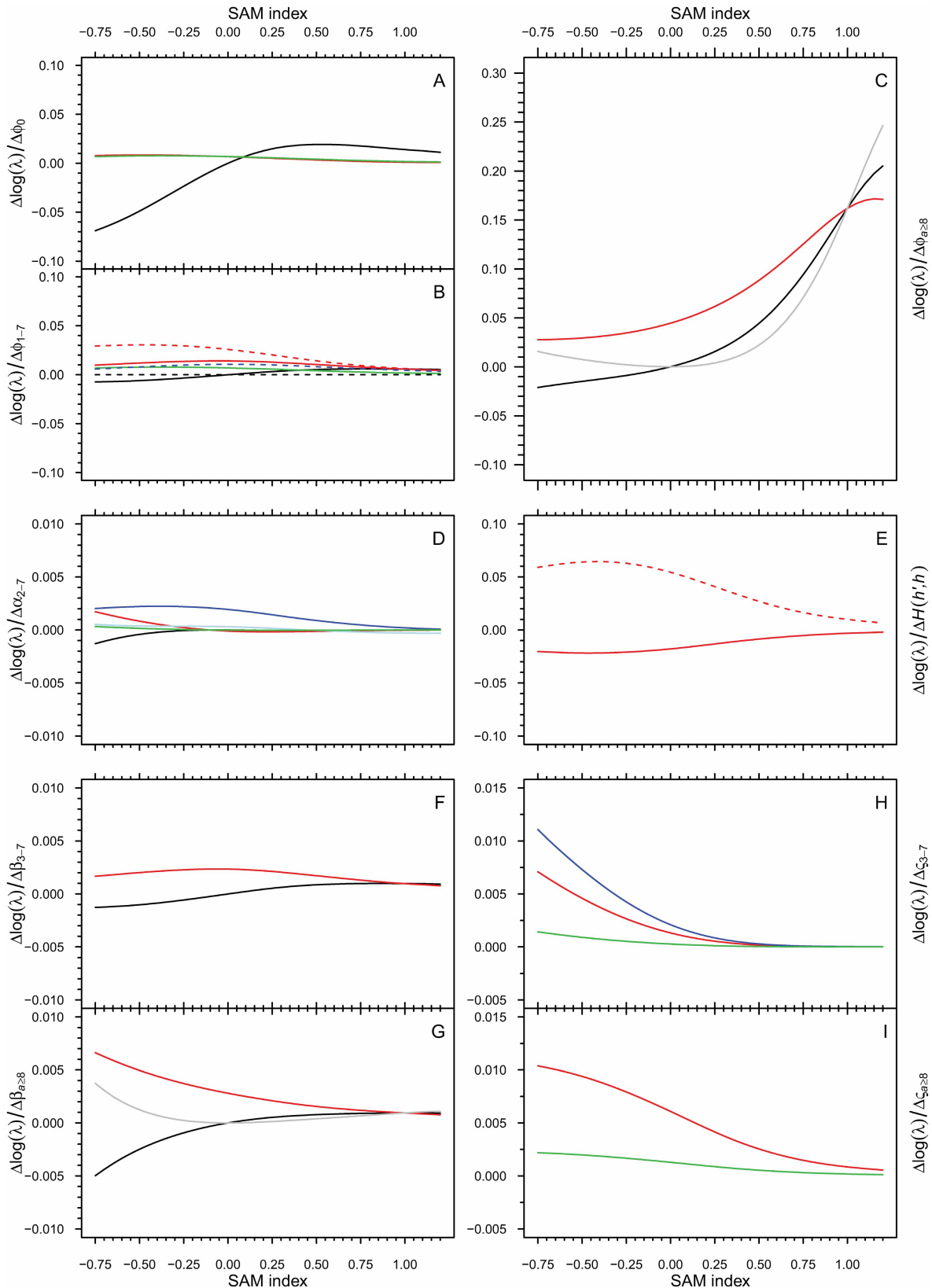
Extended Data Figure 3 | Sensitivity of the population growth rate (log(λ)) to changes in statistical parameters of vital rates describing the survival and fertility functions in integral projection models. Panels a–c show sensitivity to survival of first year females (**a**), females of ages 1 to 7 (**b**; continuous lines for pre-breeders and dashed for breeders), and ages 8 or above (**c**). Panel **d** shows sensitivity to recruitment (α) and panel **e** shows sensitivity to inheritance ($H(h'|h)$), with a dashed line for the variance intercept. Panels **f** and **g** show sensitivity to fecundity for ages 3 to 7, and 8 or above, respectively. Panels **h** and **i** show sensitivities to breeding success (ς) for ages 3 to 7, and 8 or above, respectively. Red lines are intercepts, black and grey are linear and quadratic SAM index effects, green are homozygosity weighted by locus (HL) effects, and blue are linear age effects. The scale of the vertical axes in panels **a–c** and **e** are an order of magnitude higher than in the other panels.
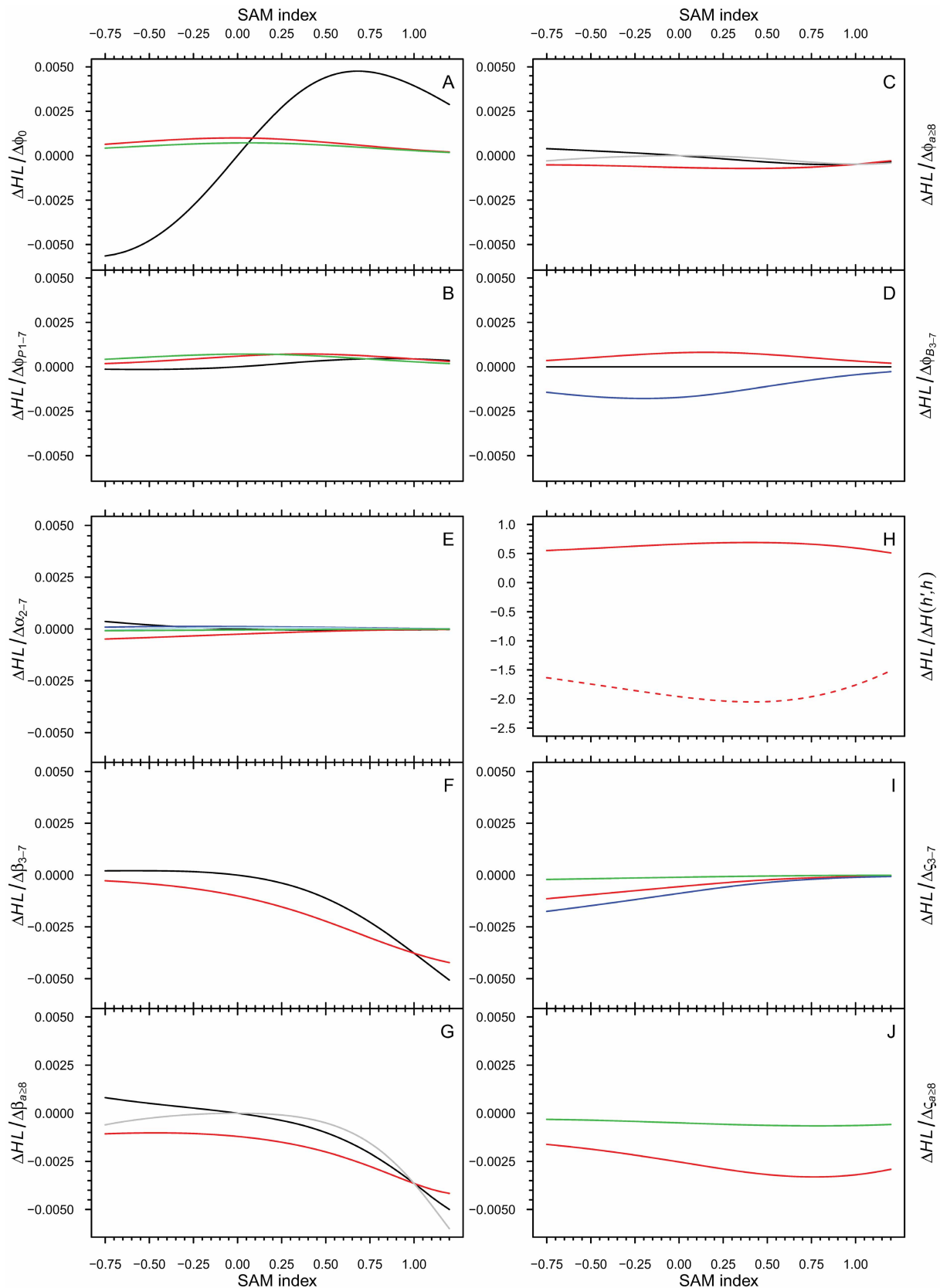
Extended Data Figure 4 | Sensitivity of the mean population homozygosity weighted by locus ($\overline{HL}$) to changes in statistical parameters of the vital rates describing the survival and fertility functions in integral projection models. Panels **a–d** show sensitivity to survival of first year females (**a**), pre-breeders of ages 1 to 7 (**b**), breeders aged 8 or above (**c**), and breeders of ages 3 to 7 (**d**). Panel **e** shows sensitivity to recruitment (α), and panels **f** and **g** show sensitivity to fecundity for ages 3 to 7, and ages 8 or above, respectively. Panel **h** shows sensitivity to inheritance ($H(h'|h)$), where the dashed line indicates the variance parameter. Panels **i** and **j** show sensitivity to breeding success for ages 3 to 7, and ages 8 or above, respectively. Colours are red for intercept, black and grey for linear and quadratic SAM index effects, respectively, green for HL effects, and blue for linear age effects.
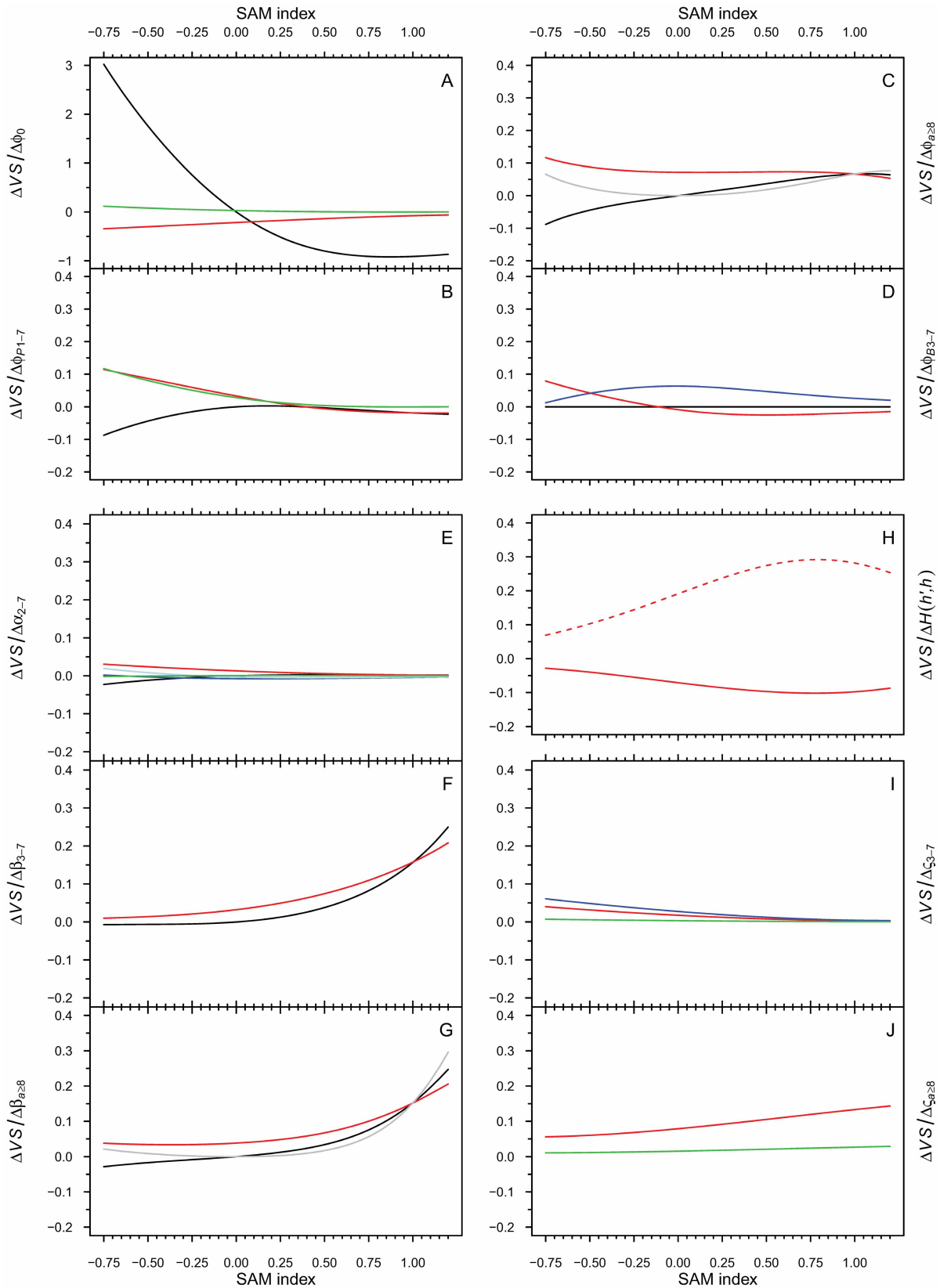
Extended Data Figure 5 | Sensitivity of the strength of viability selection (VS) on homozygosity weighted by locus (HL) to statistical parameters of vital rates describing the survival and fertility functions of integrated projection models. Panels **a–d** show sensitivity to survival of first-year females (**a**), pre-breeders of ages 1 to 7 (**b**), breeders of ages 3 to 7 (**c**) and breeders of ages 8 or above (**d**). Panel **e** shows sensitivity to recruitment (α) and panels **f** and **g** show sensitivity to fecundity for ages 3 to 7, and 8 or above, respectively.

Panel **h** shows the sensitivity to inheritance ($H(h'|h)$), where the dashed line is for the variance parameter. Panels **i** and **j** show sensitivities to breeding success for ages 3 to 7, and ages 8 or above, respectively. Colours are red for intercept, black and grey for linear and quadratic SAM index effects, respectively, green for HL effects, and dark blue and light blue for linear and quadratic age effects, respectively. Note that the scale of the vertical axis in panel **a** is an order of magnitude larger than the other panels.
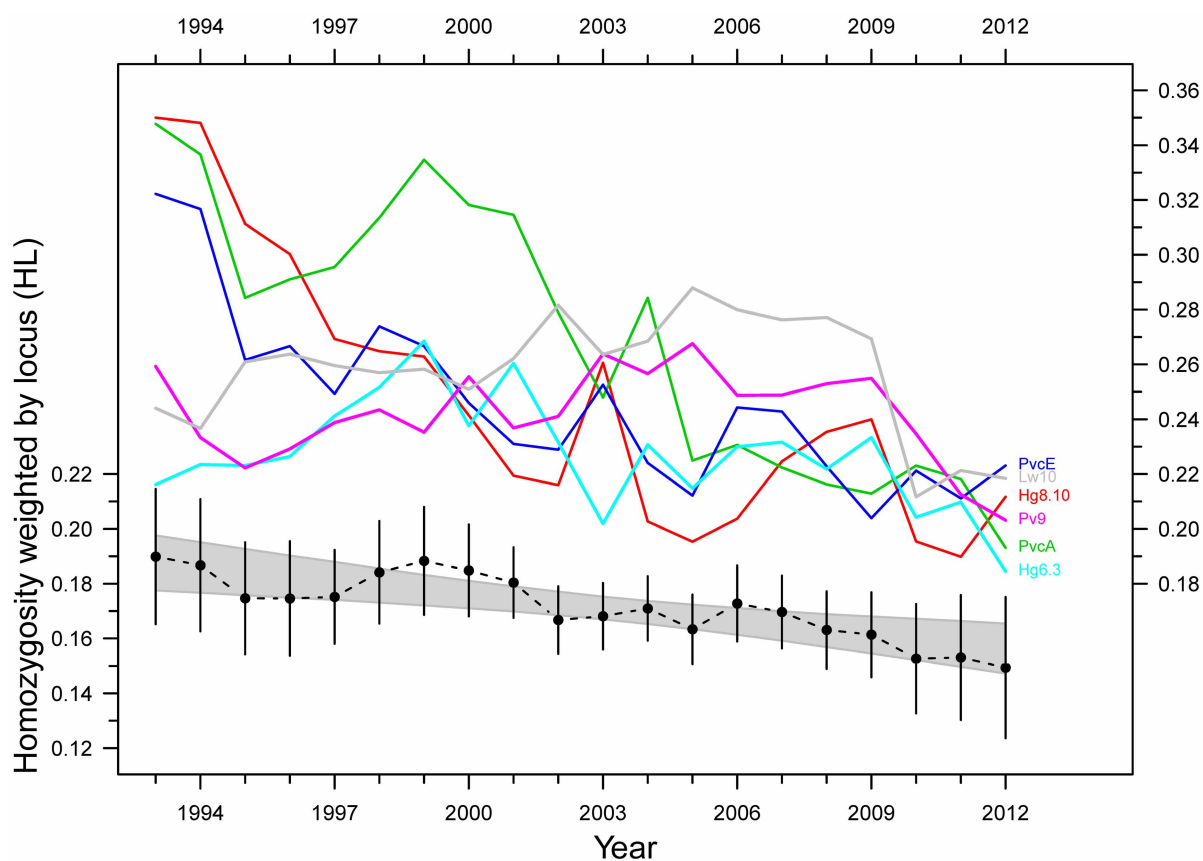
Extended Data Figure 6 | Long-term trend in breeding female homozygosity decomposed by locus. Coloured lines show mean annual homozygosity weighted by locus (HL) values for each of the loci exhibiting significant declining trends (see also Extended Data Table 3). The equivalent relationship for multilocus HL is shown below for reference, with the shaded area representing the 95% confidence interval of a fitted linear model of HL against year, and the vertical bars representing 95% confidence limits of the annual estimates.

Extended Data Table 1 | Selection of multi-event mark-recapture models of Antarctic fur seal females at Bird Island based on QAIC$_c$ (see Methods in the Supplementary Information)

Survival	Recruitment and fecundity	Breeding success	R^2_{dev}	n	deviance	QAIC$_c$	ΔQAIC$_c$
$\phi^j_{P,S=F}(A), \phi^j_N, \phi^a_{S=F,N}$	$\alpha(A^2), \beta^j_{S,F=N}, \beta^a_{S=F,N}$	$\varsigma^j_{S=F}(A), \varsigma^j_N, \varsigma^a_{S,F=N}$		28	7587.3	3628.9	125.1
$\phi^j_{P,S=F}(A+t), \phi^j_N, \phi^a_{S=F,N}(t)$	$\alpha(A^2+t), \beta^j_{S,F=N}(t), \beta^a_{S=F,N}(t)$	$\varsigma^j_{S=F}(A), \varsigma^j_N, \varsigma^a_{S,F=N}(t)$		124	7121.5	3611.1	107.3
$\phi^j_{P,S=F}(A+S), \phi^j_N, \phi^a_{S=F,N}(S^2)$	$\alpha(A^2+t), \beta^j_{S,F=N}(t), \beta^a_{S=F,N}(t)$	$\varsigma^j_{S=F}(A), \varsigma^j_N, \varsigma^a_{S,F=N}(t)$	0.87	100	7180.1	3587.2	83.4
$\phi^j_{P,S=F}(A+t), \phi^j_N, \phi^a_{S=F,N}(t)$	$\alpha(A^2+S), \beta^j_{S,F=N}(S), \beta^a_{S=F,N}(S^2)$	$\varsigma^j_{S=F}(A), \varsigma^j_N, \varsigma^a_{S,F=N}(t)$	0.95	92	7144.2	3553.2	49.4
$\phi^j_{P,S=F}(A+S), \phi^j_N, \phi^a_{S=F,N}(S^2)$	$\alpha(A^2+S), \beta^j_{S,F=N}(S), \beta^a_{S=F,N}(S^2)$	$\varsigma^j_{S=F}(A), \varsigma^j_N, \varsigma^a_{S,F=N}(t)$	0.60	68	7306.2	3531.9	28.1
$\phi^j_{P,S=F}(A+t), \phi^j_N, \phi^a_{S=F,N}(t)$	$\alpha(A^2+t), \beta^j_{S,F=N}(t), \beta^a_{S=F,N}(t)$	$\varsigma^j_{S=F}(A+H), \varsigma^j_N(H), \varsigma^a_{S,F=N}(H)$	0.98	105	7127.0	3572.8	69.0
$\phi^j_{P,S=F}(A+t+H), \phi^j_N, \phi^a_{S=F,N}(S^2)$	$\alpha(A^2+S), \beta^j_{S,F=N}(S), \beta^a_{S=F,N}(S^2)$	$\varsigma^j_{S=F}(A+H), \varsigma^j_N(H), \varsigma^a_{S,F=N}(H)$	0.82	56	7203.1	3505.5	1.7
$\phi^j_{P,S=F}(A+t), \phi^j_N, \phi^a_{S=F,N}(S^2)$	$\alpha(A^2+S+H), \beta^j_{S,F=N}(S), \beta^a_{S=F,N}(S^2)$	$\varsigma^j_{S=F}(A+H), \varsigma^j_N(H), \varsigma^a_{S,F=N}(H)$	0.82	56	7204.4	3506.1	2.3
$\phi^j_{P,S=F}(A+t+S+H), \phi^j_N(H), \phi^a_{S=F,N}(S^2)$	$\alpha(A^2+S+H), \beta^j_{S,F=N}(S), \beta^a_{S=F,N}(S^2)$	$\varsigma^j_{S=F}(A+H), \varsigma^j_N(H), \varsigma^a_{S,F=N}(H)$	0.84	58	7193.9	3505.3	1.5
$\phi^j_{P,S=F}(A+t+H), \phi^j_N(H), \phi^a_{S=F,N}(S^2)$	$\alpha(A^2+S+H), \beta^j_{S,F=N}(S), \beta^a_{S=F,N}(S^2)$	$\varsigma^j_{S=F}(A+H), \varsigma^j_N(H), \varsigma^a_{S,F=N}(H)$	0.83	56	7199.4	3503.8	0

R^2_{dev} is the amount of temporal process variance in a vital rate accounted for by the relationship with a covariate relative to the sum of sampling and total temporal process variance. Vital rates: survival (ϕ), recruitment (α), breeding propensity or fecundity (β), and breeding success (ς). Age classes (superscript): juvenile, ages 0 to 7 (j), and adult, ages 3 to 20+ (a). States (subscript): pre-breeder (P), successful breeder (S), failed breeder (F), and non-breeder (N). Variation: age-specific (A), time-specific (t), homozygosity weighted by locus (H), and Southern Annular Mode (S).

Extended Data Table 2 | Sensitivity analysis and retrospective contribution of Antarctic fur seal vital rates to λ

| Vital rates (θ_i) | Elasticities of $\log\lambda_s$ | | Variance contributions (x100) to σ_λ | | | Covariate contributions (δ) through θ_i to $\Delta\lambda$ | | | |
| | Mean $\bar{\theta}$ | Process σ_θ (x $10e^{+4}$) | variance | covariance | total | HL | | Southern Annular Mode | |
						$\%\lvert\delta_{\Delta\lambda}\rvert$	$\%\delta_{\overline{\Delta\lambda}}$	$\%\lvert\delta_{\Delta\lambda}\rvert$	$\%\delta_{\overline{\Delta\lambda}}$
ϕ_P^0	0.040	-0.31	0.0459	0.1829	0.2288	7.5	0.8 (-1.4; 3.1)	21.4	23.4 (19.6; 27.2)
ϕ_P^{1-6}	0.214	-0.30	0.0237	0.3745	0.3981	2.0	0.2 (-0.2; 0.6)	27.5	23.9 (21.1; 26.7)
$\phi_{S=F}$	0.464	-0.12	0.1332	0.2790	0.4122	–	–	27.8	30.3 (25.9; 34.7)
ϕ_N	0.104	-0.04	0.0052	0.0762	0.0814	–	–	6.0	7.0 (5.3; 8.7)
α^{3-7}	0.014	-0.01	0.0004	0.0450	0.0454	0.5	0.4 (-0.5; 1.2)	2.7	4.0 (3.4; 4.6)
$\beta_{S,F,N}^{j,a}$	0.112	-0.03	0.0053	0.1082	0.1135	–	–	7.6	7.7 (6.9; 8.5)
$\varsigma_{S,F,N}^{j,a}$	0.052	<0.01	<0.0001	0.0023	0.0023	3.2	2.4 (-2.2; 7)	–	–

Log-scaled sensitivities/elasticities of stochastic λ ($\log(\lambda_s)$) to mean and process variance of vital rates θ_i; vital rate variance contributions to temporal variation in λ ($\hat{\sigma}_\lambda$); and integrated contributions of homozygosity weighted by locus (HL) and SAM index effects on vital rates to changes in λ ($\Delta\lambda$) from 2002 to 2012, with 95% confidence intervals in parentheses. Vital rate notation is as in Extended Data Table 1.

Extended Data Table 3 | Long-term trends in heterozygosity decomposed by locus

Locus	Number of alleles	Long term population level HL trend*	Difference in heterozygosity between mother and recruited daughter[†]
Pv9	10	quadratic, $P=0.03$, Δ = - 21.7%	$\bar{x}$ = 0.15 $P<0.05$
Hg6.3	12	linear, $P=0.05$, Δ = - 22.7%	NS
Lw10	14	quadratic, $P<0.01$, Δ = - 34.7%	NS
Hg8.10	4	linear, $P<0.01$, Δ = - 18.5%	$\bar{x}$ = 0.20 $P<0.05$
Aa4	6	NS	$\bar{x}$ = 0.22 $P<0.05$
PvcA	8	linear, $P<0.01$, Δ = - 54.2%	$\bar{x}$ = 0.17 $P<0.05$
PvcE	14	linear, $P<0.01$, Δ = - 41.8%	NS
Hg1.3	12	linear, $P<0.05$, Δ = + 23.1%	NS
M11a	18	NS	NS

Details of the nine microsatellite loci are shown together with their long-term trends in homozygosity weighted by locus (HL) at the population level and average differences in HL between mothers and their recruited daughters.
* Linear or quadratic regression trends, accounting for female detectability at the study population; Δ = % change in HL over 20 years; NS non-significant (F-test for linear regression $P>0.05$).
† Paired t-test; $\bar{x}$ estimated difference; NS non-significant (F-test for linear regression $P>0.05$); $\bar{x}$ is the mean change.

LETTER

C. elegans Punctin specifies cholinergic versus GABAergic identity of postsynaptic domains

Bérangère Pinan-Lucarré[1,2]*, Haijun Tu[1,2]*, Marie Pierron[1,2], Pablo Ibáñez Cruceyra[1], Hong Zhan[1,2], Christian Stigloher[1]†, Janet E. Richmond[3] & Jean-Louis Bessereau[1,2]

Because most neurons receive thousands of synaptic inputs, the neuronal membrane is a mosaic of specialized microdomains where neurotransmitter receptors cluster in register with the corresponding presynaptic neurotransmitter release sites. In many cases the coordinated differentiation of presynaptic and postsynaptic domains implicates trans-synaptic interactions between membrane-associated proteins such as neurexins and neuroligins[1–3]. The *Caenorhabditis elegans* neuromuscular junction (NMJ) provides a genetically tractable system in which to analyse the segregation of neurotransmitter receptors, because muscle cells receive excitatory innervation from cholinergic neurons and inhibitory innervation from GABAergic neurons[4]. Here we show that *Ce-Punctin/madd-4* (ref. 5), the *C. elegans* orthologue of mammalian *punctin-1* and *punctin-2*, encodes neurally secreted isoforms that specify the excitatory or inhibitory identity of postsynaptic NMJ domains. These proteins belong to the ADAMTS (a disintegrin and metalloprotease with thrombospondin repeats)-like family, a class of extracellular matrix proteins related to the ADAM proteases but devoid of proteolytic activity[6]. *Ce-Punctin* deletion causes the redistribution of synaptic acetylcholine and GABA$_A$ (γ-aminobutyric acid type A) receptors into extrasynaptic clusters, whereas neuronal presynaptic boutons remain unaltered. Alternative promoters generate different Ce-Punctin isoforms with distinct functions. A short isoform is expressed by cholinergic and GABAergic motoneurons and localizes to excitatory and inhibitory NMJs, whereas long isoforms are expressed exclusively by cholinergic motoneurons and are confined to cholinergic NMJs. The differential expression of these isoforms controls the congruence between presynaptic and postsynaptic domains: specific disruption of the short isoform relocalizes GABA$_A$ receptors from GABAergic to cholinergic synapses, whereas expression of a long isoform in GABAergic neurons recruits acetylcholine receptors to GABAergic NMJs. These results identify Ce-Punctin as a previously unknown synaptic organizer and show that presynaptic and postsynaptic domain identities can be genetically uncoupled *in vivo*. Because human *punctin-2* was identified as a candidate gene for schizophrenia[7], ADAMTS-like proteins may also control synapse organization in the mammalian central nervous system.

In *C. elegans* each body-wall muscle cell receives excitatory and inhibitory inputs from cholinergic and GABAergic motoneurons, respectively[4]. Acetylcholine (ACh) and GABA ionotropic receptors cluster in distinct subcellular domains opposing the corresponding neurotransmitter release sites. Although the molecular mechanisms for GABA$_A$ receptor (GABA$_A$R) clustering at GABAergic NMJs are not known, genetic screens have identified three proteins expressed by muscle cells that are necessary for the synaptic clustering of heteromeric levamisole-sensitive acetylcholine receptors (L-AChR)[8–11]. LEV-10 is a transmembrane protein with a large extracellular region containing five CUB (complement, urchin epidermal growth factor, bone morphogenetic protein) domains[12]. LEV-9 and OIG-4 are secreted proteins containing CCP (complement control protein) and immunoglobulin domains, respectively[13,14]. These proteins

form an extracellular scaffold in the synaptic cleft that physically interacts with L-AChRs. Neither L-AChR, LEV-9 nor LEV-10 is localized at the synapse by itself, suggesting that additional determinants specify the synaptic localization of this molecular complex.

To identify the synaptic organizers controlling the formation of excitatory synapses, we performed a visual screen for abnormal L-AChR distribution with the use of a knock-in strain in which the red fluorescent protein TagRFP was introduced into the genomic locus of the essential L-AChR subunit *unc-29* (refs 15, 16). In this strain, UNC-29–RFP accurately localizes at cholinergic NMJs, and synaptic L-AChR clusters can be observed *in vivo* by epifluorescence (Fig. 1a). We isolated a mutant allele of the *Ce-Punctin/madd-4* gene in which UNC-29–RFP fluorescence was decreased at NMJs, and clusters were formed in extrasynaptic regions of the muscle cells (Fig. 1b and Extended Data Fig. 1a)[5]. This gene encodes three isoforms of an ADAMTS-like protein orthologous to the poorly characterized mammalian punctin-1/ADAMTS-like1 and punctin-2/ADAMTS-like3 proteins (Extended Data Fig. 1d, e)[17,18]. ADAMTS-like proteins are secreted modular proteins of the extracellular matrix that contain domains similar to ADAMTS proteases but lack the metalloprotease and disintegrin amino-terminal domains characteristic of ADAM proteases[6]. MADD-4A and MADD-4C are the longest isoforms, differing by alternative splicing of a two-residue encoding exon (Extended Data Fig. 1a, b). MADD-4B is a shorter isoform corresponding to the carboxy-terminal part of MADD-4A and MADD-4C and probably generated by the use of an alternative promoter. In *C. elegans*, MADD-4B is secreted by motoneurons and serves as a guidance cue for the growth of the so-called muscle arms[5]. These muscle cell extensions grow during post-embryonic development to reach motoneuron axons and establish *en passant* synapses at the dorsal and ventral nerve cords (Fig. 1a)[19]. Each muscle arm receives both cholinergic and GABAergic inputs.

We generated a molecular null allele of *madd-4*, namely *kr270*, by deleting the entire coding sequence of *madd-4* (Extended Data Fig. 1a). By immunofluorescence staining we found that cholinergic boutons were properly formed in this mutant, and ultrastructural analysis revealed no defect of presynaptic differentiations or synaptic cleft morphology (Fig. 1c–h and Extended Data Figs 2a–f and 3). However, L-AChR aggregates were redistributed to extrasynaptic areas (Fig. 1c–h and Extended Data Fig. 2a–f). To test whether L-AChRs were properly inserted in the plasma membrane, we first stained surface L-AChRs *in vivo* by using a strain expressing L-AChRs carrying an extracellular Myc tag[20,21]. Extrasynaptic L-AChRs were readily detected by fluorescent anti-Myc antibodies injected into the pseudocoelomic cavity of *madd-4* mutants (Extended Data Fig. 2g, h). Second, we performed an electrophysiological analysis. Pressure ejection of levamisole onto voltage-clamped muscle cells elicited similar currents in the wild type and in *madd-4* mutants (Fig. 1i), indicating that L-AChRs were fully functional. However, evoked currents triggered by motoneuron stimulation were strongly reduced (Fig. 1j), in agreement with the reduction of synaptic L-AChRs

[1]Institute of Biology, École Normale Supérieure, 75005 Paris, France. [2]University Claude Bernard Lyon 1, CGphiMC UMR CNRS 5534, 69622 Villeurbanne, France. [3]Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60607, USA. †Present address: University of Würzburg, Biocenter, Division of Electron Microscopy, Am Hubland, 97074 Würzburg, Germany.
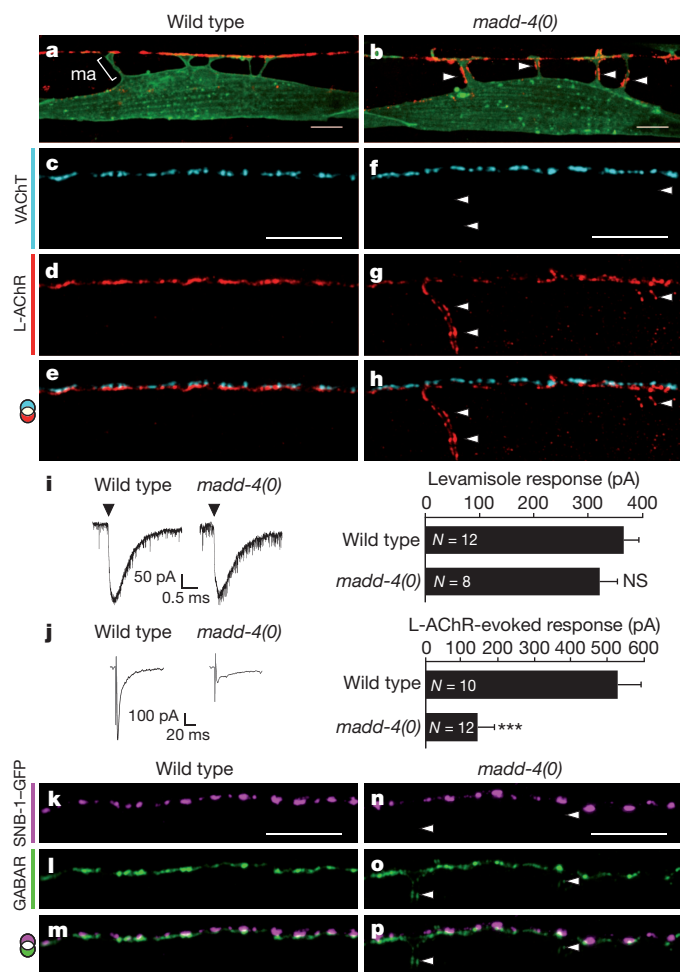*These authors contributed equally to this work.

Figure 1 | *madd-4* null mutations disrupt synaptic receptor clustering at cholinergic and GABAergic NMJs. a, b, L-AChRs (*unc-29–RFP* knock-in, red) cluster at NMJs at the nerve cords in the wild type (**a**) and at non-synaptic areas of muscle cells expressing a membrane yellow fluorescent protein (YFP; green) in a *madd-4(kr249)* mutant (**b**). The bracket indicates one muscle arm (ma). **c–h,** Immunofluorescence staining of dorsal nerve cords shows that cholinergic boutons labelled with an anti-vesicular ACh transporter (VAChT)/ UNC-17 antibody (cyan) (**c, f**) are similar in the wild type and in the *madd-4(kr270)* null mutant, whereas L-AChR clusters stained with an anti-UNC-38 antibody (red) (**d, g**) redistribute away from acetylcholine release sites in a *madd-4(kr270)* mutant. **i,** Peak currents evoked by a 100 ms pressure application of 500 µM levamisole were recorded from voltage-clamped body-wall muscles and were similar in the wild type and in the *madd-4(kr270)* mutant. Arrowheads indicate the application onset. **j,** Isolated L-AChR-dependent currents evoked by motoneuron depolarization are reduced in *madd-4(kr270)* null mutants compared with those in wild-type animals. **k–p,** GABAergic boutons revealed by anti-GFP immunofluorescence of transgenics expressing the vesicle-associated membrane protein (VAMP)/ SNB-1–GFP in GABA motoneurons (magenta) (**k, n**) form properly in a *madd-4(kr270)* mutant, whereas GABA$_A$Rs stained by anti-UNC-49 antibody (green) (**l, o**) redistribute at extrasynaptic sites in the *madd-4(kr270)* mutant (**n–p**) in contrast with the wild type (**k–m**). In this figure, white arrowheads indicate extrasynaptic clusters of receptors. **e, h, m, p,** Channel overlays along with a colour mixing guide. Scale bars, 10 µm. Data are presented as means and s.e.m.; N = number of worms; Mann–Whitney test: ***$P \leq 0.0001$; NS, not significant.

detected by UNC-29–RFP visualization and UNC-38 immunofluorescence staining. In *C. elegans*, cholinergic NMJs also contain homomeric N-AChRs (nicotine-sensitive AChRs) composed of the ACR-16 subunit[22,23]. The distribution of N-AChR was similar to that of L-AChRs in *madd-4* mutants (Extended Data Fig. 2i, j). In addition, the muscle GABA$_A$R UNC-49 formed clusters that redistributed at extrasynaptic

areas of muscle cells, in a similar manner to AChRs, whereas GABAergic boutons were properly formed in *madd-4* mutants (Fig. 1k–p)[24]. Therefore *madd-4* is required for the differentiation of postsynaptic domains independently from presynaptic differentiation at cholinergic and GABAergic NMJs.

We previously demonstrated that the removal of either LEV-9 or LEV-10 scaffold protein leads to the complete disruption of the L-AChR clustering complex[12,13]. Because extrasynaptic L-AChRs still formed clusters in *madd-4* mutants, we tested whether these clusters remained associated with the scaffolding machinery. First, we observed that LEV-9 and LEV-10 were present in synaptic and extrasynaptic L-AChR clusters in *madd-4* mutants (Extended Data Fig. 2k–p). Second, we genetically inactivated *lev-9* and *lev-10* in *madd-4* mutants and observed that L-AChRs were no longer clustered (Extended Data Fig. 2q–s). Hence MADD-4 is not required for the assembly of the L-AChR clustering machinery itself but is necessary to recruit or stabilize L-AChR complexes at postsynaptic sites.

To analyse the subcellular distribution of MADD-4 variants, we generated functional green fluorescent protein (GFP)-tagged reporters for MADD-4L long isoforms (which globally refers to MADD-4A and MADD-4C) and MADD-4B in the context of a rescuing fosmid construct (Extended Data Figs 1c and 4)[25]. First, we verified that MADD-4 isoforms were indeed secreted proteins as suggested by the presence of predicted signal peptides (Extended Data Figs 1d and 5a–f). Second, we observed that MADD-4L–GFP was juxtaposed to cholinergic terminals but not to GABAergic terminals, whereas MADD-4B–GFP was juxtaposed to both types of bouton (Fig. 2). These observations are consistent with the differential expression of *madd-4L* and *madd-4B* promoter fragments in excitatory motoneurons or in both types of motoneuron, respectively[5]. Third, we injected anti-GFP antibodies *in vivo* and specifically stained extracellular MADD-4L–GFP and MADD-4B–GFP at NMJs, indicating the secretion and retention of these proteins in the synaptic cleft (Extended Data Fig. 5g–l). Because MADD-4B was previously reported to require the kinesin UNC-104 to be properly secreted at the dorsal cord[5], we analysed L-AChR and GABA$_A$R distribution in *unc-104* mutants and observed extrasynaptic receptor clusters as in *madd-4* mutants (Extended Data Fig. 5m–p). To test whether the localization of MADD-4A–GFP and MADD-4B–GFP depended on AChRs, we analysed their distribution in an *unc-29; acr-16* double mutant and found that they were properly localized at cholinergic NMJs (Extended Data Fig. 6a–d). In addition, MADD-4B–GFP was retained at GABAergic synapses in the absence of GABA$_A$R (Extended Data Fig. 6e–h). MADD-4 isoforms are therefore secreted and retained at synapses independently from postsynaptic receptors, and probably serve as molecular landmarks to define the synaptic zone on the postsynaptic membrane.

Because our previous observations were made in adult nematodes, the redistribution of L-AChR and GABA$_A$R at extrasynaptic areas might have resulted from defective formation or defective maintenance of receptor clustering. Soon after hatching, clusters of L-AChRs and GABA$_A$Rs spread outside the nerve cords in *madd-4* mutant, in contrast with the clusters confined to NMJs in the wild type (Fig. 3a–h), indicating that *madd-4* may be required early during synapse formation. Consistently, MADD-4L–GFP and MADD-4B–GFP were readily detected at the nerve cords at this stage (Fig. 3i, j). To test directly whether MADD-4 can induce receptor clustering, we expressed MADD-4A–GFP ectopically in muscle cells of wild-type animals. Large ectopic L-AChR clusters co-localized with ectopic MADD-4A–GFP punctae, whereas the intensity of synaptic clusters was greatly decreased, suggesting that MADD-4A overexpression at extrasynaptic sites outcompeted the normal complement of receptors at the synapse (Fig. 3k–m). These ectopic clusters could also form in the absence of MADD-4B (data not shown). In addition, ectopic expression of MADD-4A–GFP in the FLP sensory neurons triggered ectopic clustering of L-AChRs on the internal side of the muscle close to the anterior FLP process (Fig. 3n–p). Altogether, these data suggest that *madd-4* has an early function during synaptogenesis and is able to instruct postsynaptic receptor clustering at the synapse.
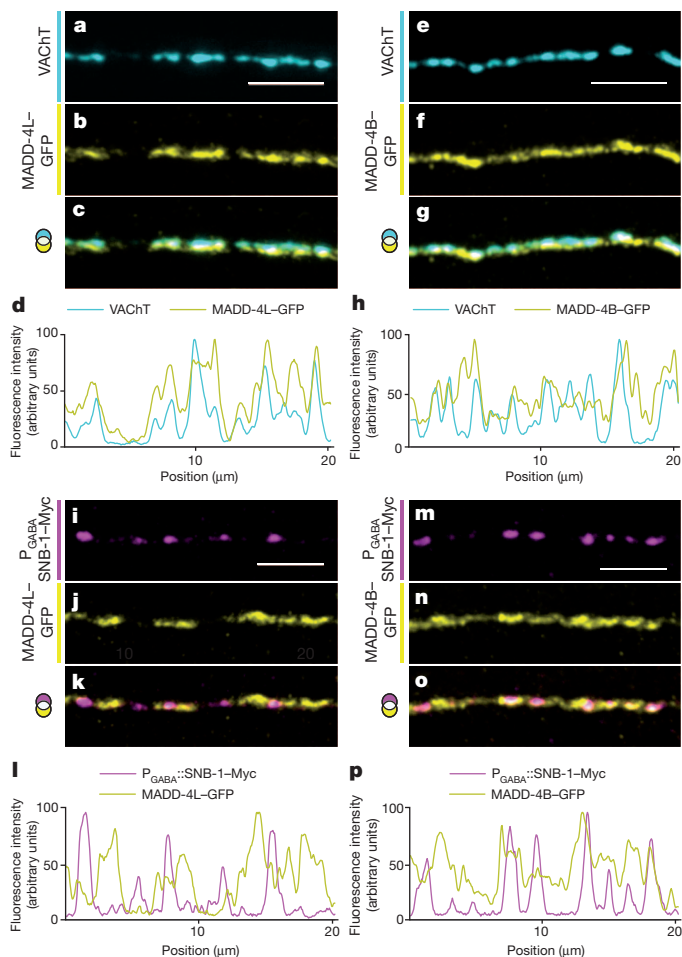
Figure 2 | The long isoform of MADD-4 localizes at cholinergic NMJs, whereas the short isoform is at cholinergic and GABAergic NMJs.
a–h, Immunofluorescence staining indicates that both MADD-4L–GFP (*krEx1067*) (**b**) and MADD-4B–GFP (*krEx1069*) (**f**) are juxtaposed to cholinergic boutons (anti-VAChT/UNC-17 antibody) (**a**, **e**). **d**, **h**, Fluorescence profile of GFP (yellow) and UNC-17 (cyan) along the cord from pictures above. **i–p**, Visualization of GABA boutons (**i**, **m**) shows that MADD-4L–GFP (**j**) is not present at GABA synapses, unlike MADD-4B–GFP (**n**). **l**, **p**, Fluorescence profile of GFP (yellow) and VAMP/SNB-1-MYC (magenta) in GABA boutons. **c**, **g**, **k**, **o**, Channel overlays along with a colour mixing guide. Scale bars, 5 μm.

Given the differential localization of the short and long isoforms of MADD-4, we speculated that each isoform might have a specific function in the differentiation of excitatory and inhibitory postsynapses. We individually reintroduced each isoform driven by its own promoter in a *madd-4* null mutant. Synaptic L-AChR content, estimated by measuring UNC-29–RFP fluorescence at the dorsal nerve cord, was reduced by 60% in a *madd-4* null mutant. This decrease was fully rescued by MADD-4L expression, whereas MADD-4B had no rescuing activity (Fig. 4a). To analyse the contribution of each isoform further, we used mutations that specifically disrupt either the short or the long isoforms (referred to as *madd-4B(0)* and *madd-4L(0)*, respectively; Fig. 4b and Extended Data Fig. 1a). Electrophysiological analysis indicated that the total amount of L-AChRs present at the muscle cell surface was similar to that in the wild type in both mutants (Fig. 4c). In contrast, L-AChR-dependent evoked currents were significantly reduced in *madd-4L(0)* mutants but were wild-type in *madd-4B(0)* mutants (Fig. 4d). Consistently, L-AChR localization was wild-type in *madd-4B(0)* mutants or when MADD-4L was expressed in the *madd-4* null mutant (Extended Data Figs 7 and 4d–i, respectively). The two long isoforms A and C had similar L-AChR clustering activities (Extended Data Fig. 8a). Taken
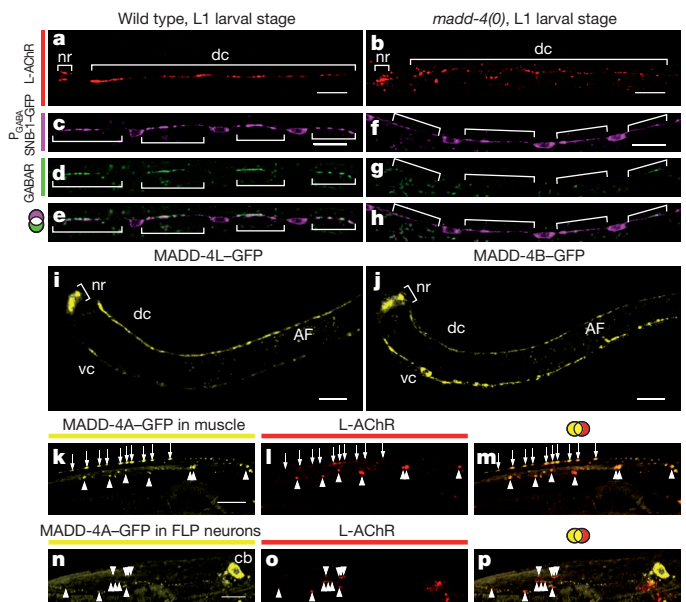

Figure 3 | MADD-4 instructs receptor clustering. a, **b**, L-AChR clusters revealed at the L1 larval stage in the *unc-29–RFP* knock-in are disorganized in a *madd-4(kr249)* mutant (**b**), in contrast with the wild type (**a**). Brackets indicate the nerve ring (nr) and the dorsal cord (dc). **c–h**, At the ventral cord of *madd-4(kr270)* mutant L1 larvae (**f–h**), GABA$_A$Rs (*krSi2[P$_{unc-49}$::unc-49-RFP]*) (**d**, **g**) are not clustered at synapses although GABAergic presynaptic sites (*juIs1[P$_{unc-25}$::snb-1–GFP]*) (**c**, **f**) are correctly formed, in contrast with the wild type (**c–e**). Brackets indicate the position of GABAergic boutons, and the cell bodies of GABA motoneurons are visible between the brackets. **i**, **j**, At the L1 larval stage, MADD-4L–GFP (**i**) is mainly localized at the nerve ring and at the dorsal cord, whereas MADD-4B–GFP (**j**) is localized at the nerve ring and at both the ventral (vc) and dorsal cords. AF, non-specific autofluorescence of intestinal granules. **k–m**, In adult nematodes, ectopic expression of the long isoform MADD-4A–GFP in muscle cells (**k**) causes ectopic clustering of L-AChRs (*unc-29–RFP*) (**l**). **n–p**, Ectopic expression of MADD-4A–GFP in the FLP sensory neurons (**n**) causes ectopic clustering of L-AChRs on the internal side of the muscle (*unc-29–RFP*) (**o**) in *madd-4(kr249)* mutants. cb, FLP cell body. Arrows indicate the position of SAB synapses; arrowheads show ectopic clustering. **e**, **h**, **m**, **p**, Channel overlays along with a colour mixing guide. Scale bars, 10 μm.

together, these results demonstrate that the long isoforms are necessary and sufficient for localizing L-AChR clusters at NMJs.

Because MADD-4B is required for the guidance of postsynaptic muscle arms[5], the mislocalization of L-AChRs in *madd-4* mutants might have been a secondary consequence of development defects of the muscle arms. However, several lines of evidence demonstrate that MADD-4 controls L-AChR synaptic clustering independently of muscle arm growth. First, disruption of *madd-4* causes the disappearance of L-AChR clusters at SAB synapses (Extended Data Fig. 2a–f), which are directly made *en passant* between SAB axons and muscle cell soma with no requirement for muscle arm formation[4]. Second, the *madd-4L(0)* mutation causes L-AChR localization defects as severe as those in the *madd-4* null mutant, whereas muscle arms are wild-type (Extended Data Fig. 9). Third, the *madd-4B(0)* mutant is impaired in muscle arm development[5] but it does not show defects in L-AChR localization or function (Fig. 4c, d and Extended Data Fig. 7a–f).

We then analysed the localization of GABA$_A$Rs in the absence of MADD-4 isoforms. Specific ablation of the short isoform caused a striking phenotype: GABA$_A$Rs still assembled into synaptic clusters but were now present at cholinergic synapses (Fig. 5a–c, g–i). A similar redistribution was achieved by re-expressing MADD-4L using its own promoter or a cholinergic-specific promoter in the *madd-4* null mutant (Extended Data Figs 4m–o and 8b–f). In contrast, GABA$_A$Rs were properly clustered at GABAergic NMJs in *madd-4L(0)* mutants (Fig. 5d–f, j–l) or in transgenic lines expressing MADD-4B in *madd-4* null mutants
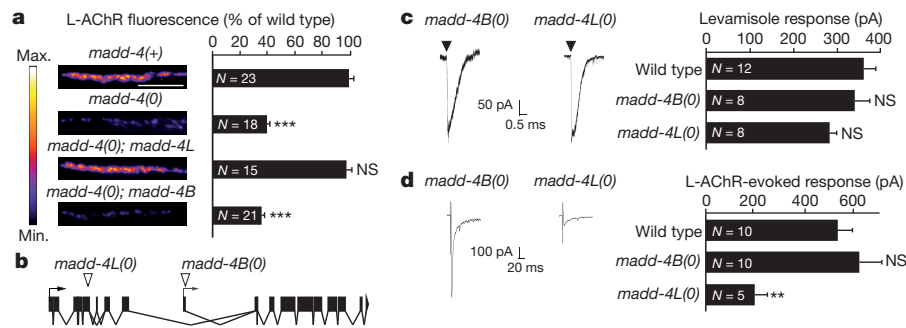
Figure 4 | Long MADD-4 isoforms ensure synaptic clustering of L-AChR.
a, Fluorescence intensity of the L-AChR subunit UNC-29–RFP at the dorsal cord of the wild type and of *madd-4* null mutants rescued with fosmids expressing either the long (MADD-4L) or short (MADD-4B) isoforms. Scale bar, 10 μm. **b**, Genomic organization of *madd-4* (black boxes represent exons), and the position of two mutations (open arrowheads) that specifically disrupt *madd-4L* (*madd-4(ttTi103747)*, referred to as *madd-4L(0)*) or *madd-4B* (*madd-4(tr185)*, referred to as *madd-4B(0)*). Arrows indicate the beginning of the open reading frame of *madd-4L* (black) and *madd-4B* (grey). **c**, Peak currents evoked by 100 ms pressure application of levamisole in *madd-4B(0)* and *madd-4L(0)* mutants. **d**, Isolated L-AChR-evoked currents are not impaired in *madd-4B(0)* mutant but are strongly reduced in the *madd-4L(0)* mutant (Mann–Whitney test indicated no significant difference from the *madd-4(kr270)* null mutant presented in Fig. 1j). Results are presented as means and s.e.m.; $N =$ number of worms; Student's t-test (**a**) or Mann–Whitney test (**c**, **d**): ***$P \leq 0.0001$; **$P \leq 0.005$; NS, not significant.
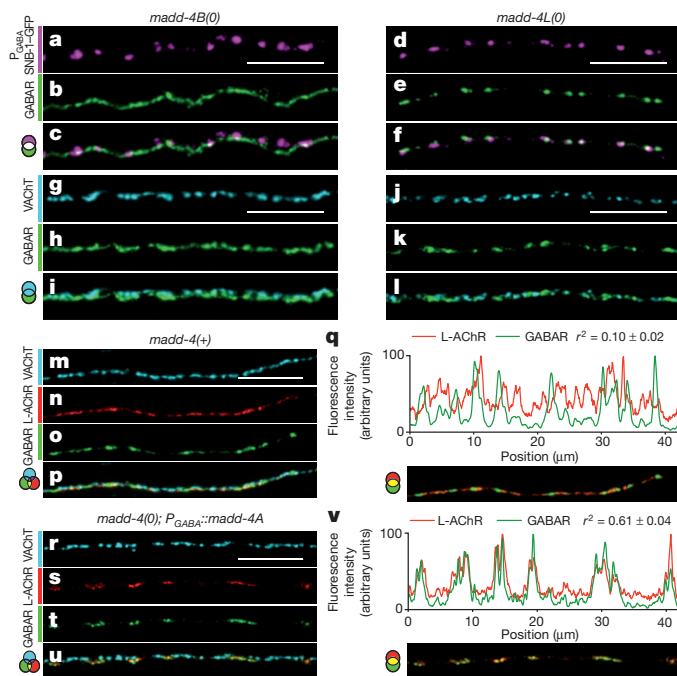

Figure 5 | Differential control of postsynaptic identity by the long and short MADD-4 isoforms. a–l, Immunofluorescence staining of GABA$_A$R (anti-UNC-49) (**b**, **e**, **h**, **k**), GABA boutons (P_{unc-25}::*snb-1*–GFP) (**a**, **d**) and ACh boutons (anti-UNC-17/VAChT antibody) (**g**, **j**) reveals that GABA$_A$Rs redistribute opposite cholinergic terminals in the *madd-4B(0)* mutant (**a–c**, **g–i**), whereas they are properly localized in the *madd-4L(0)* mutant (**d–f**, **j–l**). **m–v**, Simultaneous detection of cholinergic terminals (anti-UNC-17 immunofluorescence) (**m**, **r**), L-AChRs (UNC-29–RFP) (**n**, **s**) and GABA$_A$Rs (anti-UNC-49 immunofluorescence) (**o**, **t**) in *madd-4(+)* and in *madd-4(kr249)* null mutants expressing the long isoform, MADD-4A, in GABA motoneurons [*madd-4(0)*; p_{GABA}::*madd-4A*]. The fluorescence profiles of L-AChRs (red) and GABA$_A$Rs (green) along the cord are shown (**q**, **v**). Co-localization of L-AChRs and GABA$_A$Rs was quantified in 23 and 19 worms, respectively. r^2 values are the means of the squares of Pearson's coefficient (mean ± s.e.m.). A value close to 0 indicates a lack of co-localization; a value close to 1 indicates complete co-localization because all calculated r values were positive. The r^2 values of the two genotypes are statistically different by analysis of variance. **c**, **f**, **i**, **l**, **p**, **u**, Channel overlays along with a colour mixing guide. Scale bars, 10 μm.

(Extended Data Fig. 4p–r). These striking results indicate, first, that the short isoform is essential for clustering GABA$_A$Rs at GABAergic NMJs and second, that the long isoform might be able to cluster both AChRs and GABA$_A$Rs irrespective of presynaptic identity. To test this latter hypothesis, we expressed MADD-4A in GABAergic motoneurons. L-AChRs were now present opposite GABAergic release sites and mixed with GABA$_A$Rs as assessed by the strong co-localization of the two types of receptor (Fig. 5m–v and Extended Data Fig. 8l–n). These results raise an apparent paradox. Indeed, in wild-type animals the long isoforms are present only at cholinergic NMJs, yet GABA$_A$Rs are excluded from these synapses. A simple model would propose that an interaction between the short and long isoforms at cholinergic junctions could inhibit the GABA$_A$R clustering activity of the long isoforms. In agreement with this model, less GABA$_A$Rs co-localized with L-AChRs at cholinergic NMJs when co-expressing both MADD-4A and MADD-4B in cholinergic neurons than when expressing the long isoform alone (Extended Data Fig. 8g–k, l–n). To test whether long and short isoforms can interact molecularly, we performed co-immunoprecipitation experiments in human embryonic kidney (HEK) cells. Results showed that MADD-4 isoforms can undergo homophilic and heterophilic interactions (Extended Data Fig. 10), which might account for the functional interactions observed *in vivo*.

We have identified the ADAMTS-like *Ce-Punctin* gene as a critical organizer of postsynaptic domains at *C. elegans* neuromuscular junctions. The short and long Ce-Punctin isoforms are differentially expressed by cholinergic and GABAergic motoneurons. Long Ce-Punctin isoforms localize AChR clusters at excitatory NMJs, whereas the short Ce-Punctin isoform prevents the recruitment of GABARs to cholinergic synapses and promotes GABAR clustering at GABA synapses. In contrast with models of synaptic organization relying on trans-synaptic interactions of proteins that ultimately connect presynaptic and postsynaptic membranes[1–3], our results suggest an alternative model that relies on the local release of a neurally secreted protein that acts as a molecular landmark triggering the clustering of specific receptors. This might be achieved through the activation of a signal-transducing receptor, as in the agrin/MuSK pathway at the mammalian NMJ[26,27]. Alternatively, the long Ce-Punctin isoforms might directly recruit the L-AChR-associated scaffold through protein–protein interactions, because ADAMTS-like proteins are known to bind extracellular proteins such as fibrillin and latent transforming growth factor-β binding protein[28,29]. The ability of ADAMTS-like proteins to organize the synapse was unknown until we characterized the phenotype of the *Ce-punctin/madd-4* mutants. Because the *ADAMTSL3/punctin-2* gene is expressed in the central nervous system (based on Allen Mouse Brain Atlas[30]) and was identified as a candidate

gene for schizophrenia[7], it will be interesting to test whether the synaptic functions of Ce-Punctin have been conserved during evolution.

METHODS SUMMARY

Transgenic strains were generated by microinjection into the gonad of adult hermaphrodites, using standard techniques. We performed genome engineering by homologous recombination techniques to create deletion and single-copy insertion alleles[16]. Plasmids were generated by using standard molecular biology methods, and all sequences are available from the authors on request. We performed fosmid recombineering as described previously[25].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Sudhof, T. C. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* **455**, 903–911 (2008).
2. Shen, K. & Scheiffele, P. Genetics and cell biology of building specific synaptic connectivity. *Annu. Rev. Neurosci.* **33**, 473–507 (2010).
3. Siddiqui, T. J. & Craig, A. M. Synaptic organizing complexes. *Curr. Opin. Neurobiol.* **21**, 132–143 (2011).
4. White, J. G., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
5. Seetharaman, A. *et al.* MADD-4 is a secreted cue required for midline-oriented guidance in *Caenorhabditis elegans*. *Dev. Cell* **21**, 669–680 (2011).
6. Apte, S. S. A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. *J. Biol. Chem.* **284**, 31493–31497 (2009).
7. Dow, D. J. *et al.* ADAMTSL3 as a candidate gene for schizophrenia: gene sequencing and ultra-high density association analysis by imputation. *Schizophr. Res.* **127**, 28–34 (2011).
8. Fleming, J. T. *et al. Caenorhabditis elegans* levamisole resistance genes *lev-1, unc-29,* and *unc-38* encode functional nicotinic acetylcholine receptor subunits. *J. Neurosci.* **17**, 5843–5857 (1997).
9. Boulin, T. *et al.* Eight genes are required for functional reconstitution of the *Caenorhabditis elegans* levamisole-sensitive acetylcholine receptor. *Proc. Natl Acad. Sci. USA* **105**, 18590–18595 (2008).
10. Towers, P. R., Edwards, B., Richmond, J. E. & Sattelle, D. B. The *Caenorhabditis elegans lev-8* gene encodes a novel type of nicotinic acetylcholine receptor alpha subunit. *J. Neurochem.* **93**, 1–9 (2005).
11. Culetto, E. *et al.* The *Caenorhabditis elegans unc-63* gene encodes a levamisole-sensitive nicotinic acetylcholine receptor alpha subunit. *J. Biol. Chem.* **279**, 42476–42483 (2004).
12. Gally, C., Eimer, S., Richmond, J. E. & Bessereau, J. L. A transmembrane protein required for acetylcholine receptor clustering in *Caenorhabditis elegans*. *Nature* **431**, 578–582 (2004).
13. Gendrel, M., Rapti, G., Richmond, J. E. & Bessereau, J. L. A secreted complement-control-related protein ensures acetylcholine receptor clustering. *Nature* **461**, 992–996 (2009).
14. Rapti, G., Richmond, J. & Bessereau, J. L. A single immunoglobulin-domain protein required for clustering acetylcholine receptors in *C. elegans*. *EMBO J.* **30**, 706–718 (2011).
15. Richard, M., Boulin, T., Robert, V. J., Richmond, J. E. & Bessereau, J. L. Biosynthesis of ionotropic acetylcholine receptors requires the evolutionarily conserved ER membrane complex. *Proc. Natl Acad. Sci. USA* **110**, E1055–E1063 (2013).
16. Robert, V. & Bessereau, J. L. Targeted engineering of the *Caenorhabditis elegans* genome following Mos1-triggered chromosomal breaks. *EMBO J.* **26**, 170–183 (2007).
17. Hirohata, S. *et al.* Punctin, a novel ADAMTS-like molecule, ADAMTSL-1, in extracellular matrix. *J. Biol. Chem.* **277**, 12182–12189 (2002).
18. Hall, N. G., Klenotic, P., Anand-Apte, B. & Apte, S. S. ADAMTSL-3/punctin-2, a novel glycoprotein in extracellular matrix related to the ADAMTS family of metalloproteases. *Matrix Biol.* **22**, 501–510 (2003).
19. Dixon, S. J. & Roy, P. J. Muscle arm development in *Caenorhabditis elegans*. *Development* **132**, 3079–3092 (2005).
20. Gottschalk, A. & Schafer, W. R. Visualization of integral and peripheral cell surface proteins in live *Caenorhabditis elegans*. *J. Neurosci. Methods* **154**, 68–79 (2006).
21. Boulin, T. *et al.* Positive modulation of a Cys-loop acetylcholine receptor by an auxiliary transmembrane subunit. *Nature Neurosci.* **15**, 1374–1381 (2012).
22. Touroutine, D. *et al. acr-16* encodes an essential subunit of the levamisole-resistant nicotinic receptor at the *Caenorhabditis elegans* neuromuscular junction. *J. Biol. Chem.* **280**, 27013–27021 (2005).
23. Francis, M. M. *et al.* The Ror receptor tyrosine kinase CAM-1 is required for ACR-16-mediated synaptic transmission at the *C. elegans* neuromuscular junction. *Neuron* **46**, 581–594 (2005).
24. Bamber, B. A., Beg, A. A., Twyman, R. E. & Jorgensen, E. M. The *Caenorhabditis elegans unc-49* locus encodes multiple subunits of a heteromultimeric GABA receptor. *J. Neurosci.* **19**, 5348–5359 (1999).
25. Tursun, B., Cochella, L., Carrera, I. & Hobert, O. A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. *PLoS ONE* **4**, e4625 (2009).
26. Glass, D. J. *et al.* Agrin acts via a MuSK receptor complex. *Cell* **85**, 513–523 (1996).
27. Wu, H., Xiong, W. C. & Mei, L. To build a synapse: signaling pathways in neuromuscular junction assembly. *Development* **137**, 1017–1033 (2010).
28. Tsutsui, K. *et al.* ADAMTSL-6 is a novel extracellular matrix protein that binds to fibrillin-1 and promotes fibrillin-1 fibril formation. *J. Biol. Chem.* **285**, 4870–4882 (2010).
29. Le Goff, C. *et al.* ADAMTSL2 mutations in geleophysic dysplasia demonstrate a role for ADAMTS-like proteins in TGF-β bioavailability regulation. *Nature Genet.* **40**, 1119–1123 (2008).
30. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

Author Contributions B.P.-L., H.T., M.P. and J.-L.B. conceived the ideas and designed the experiments. B.P.-L, H.T., M.P., P. I.C., H.Z., C.S., J.E.R. and J.-L.B. performed the experiments. B.P.-L, H.T., M.P., J.E.R. and J.-L.B. analysed the data. B.P.-L. and J.-L.B. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-L.B. (jean-louis.bessereau@univ-lyon1.fr).

METHODS

***C. elegans* strains and media.** Strains were grown at 20 °C unless otherwise stated on NGM agar plates seeded with *Escherichia coli* OP50 as a food source[31]. The wild-type strain was *C. elegans* N2 Bristol. Cohorts of synchronized L1 larvae were obtained by collecting eggs at 20 °C for 1 h, transferring the eggs to 15 °C for 14 h and then to 20 °C until hatching. Freshly harvested L1 larvae were observed 2 h after hatching. Transgenic lines based on the rescue of *lin-15(n765ts)* were grown at 25 °C before immunofluorescence staining.

The following mutant alleles and transgenes were used in this study:

LGI: *unc-29(x29)*, *unc-29(kr208::tagRFP)*[15], *madd-4(kr270)*, *madd-4(kr249)*, *madd-4(ok2862)*, *madd-4(tr185)*[5], *madd-4(ttTi103747)*; LGII: *unc-104(e1265)*; LGIII: *unc-49(e407)*; LGIV: *juIs1[P_unc-25::snb-1–GFP]*[32]; LGV: *acr-16(ok789)*; LGX: *lin-15(n765)*, *lev-9(ox177)*, *lev-9(kr184::T7)*[13]; *ljEx42 [P_unc-38::unc-38-MYC-6xHIS-2× Myc, rol-6]*[20], *jaSi4[P_myo-3::acr-16-GFP]*[33], *trIs25 [pPRRF138.2(him-4p::MB::YFP)]*[5].

The following transgenic lines were created for this study:

(1) In *unc-29(kr208::tagRFP) madd-4(kr249); lin-15(n765)*: *krEx1028 [pBP31; pEKL15]; krEx1097, krEx1098 and krEx1099 [pBP8; pEKL15]; krEx1100, krEx1101 and krEx1102 [pBP11; pEKL15]; krEx1085, krEx1086 and krEx1087 [pMP9; pEKL15]; krEx1093 and krEx1094 [pBP23; pMP1; pEKL15]; krEx1039, krEx1135 and krEx1136 [pEKL15]; krEx1042, krEx1043 and krEx1044 [pBP23; pEKL15]; krEx1067 and krEx1068 [pBP35; pBP33; pEKL15]; krEx1069 and krEx1070 [pBP36; pBP33; pEKL15]; krEx1084 [pBP34; pBP33; pEKL15]; krEx1156 [pBP70; pRF4]*.

(2) In *unc-29(kr208::tagRFP); lin-15(n765)*: *krEx970 [P_him-4::mb-yfp; pEKL15]*.

Generation of deletion and single-copy insertion alleles by homologous recombination. The *madd-4(kr270)* molecular null allele was generated by using the *Mos1*-mediated deletion (*Mos*DEL) technique as described[16,34]. The entire coding sequence of *madd-4*, starting from the ATG of *madd-4A* and *madd-4C*, was replaced with the *Caenorhabditis briggsae unc-119* wild-type gene. The pBP6 vector was injected as a rescue template into a strain containing the *ttTi21487 Mos1* insertion, internal to the *madd-4* gene. In pBP6, a 3-kilobase (kb) left *madd-4* homology sequence and a 2-kb right homology sequence flank the *unc-119* marker.

The *krSi4* allele encodes a *madd-4A–GFP* minigene and was generated by using the *Mos1*-mediated single-copy insertion (*Mos*SCI) technique as described[16,35]. The *madd-4A–GFP* minigene is composed of 2,652 base pairs (bp) of *madd-4A* promoter, a genomic DNA fragment containing the first 5 exons of *madd-4A*, a complementary DNA fragment corresponding to the last 12 exons, and the sequence of eGFP and the 3′ untranslated region (UTR) sequence of *unc-54*. The EG4322 strain was injected with pBP27 as a rescue template, a vector in which the *madd-4A–GFP* minigene was inserted into pCFJ151 by using XhoI and SdaI sites.

The *krSi2* allele was generated by *Mos*SCI. It encodes a *unc-49B-tagRFP* minigene, composed of 4,070 bp of *unc-49* promoter, a genomic DNA fragment containing the first four exons of *unc-49*, a cDNA fragment corresponding to the last eight exons of *unc-49B* isoform and the 3′ UTR sequence of *unc-54*. The tagRFP sequence was inserted in the TM3–TM4 intracellular loop of the UNC-49B subunit. The EG4322 strain was injected with the pTH22.1 vector, in which the *unc-49B-tagRFP* minigene fused with the 3′ UTR sequence of *unc-54* was inserted into pCFJ151 by using SbfI and SpeI sites.

Germline transformation. Transformation was performed by microinjection of plasmid DNA into the gonad as described[36]. For generation of the fosmid-based lines, the pBP35 fosmid (or pBP36) was injected at 10 ng µl⁻¹ with pBP33 (*P_unc-47::snb-1–GFP*) at 30 ng µl⁻¹, pEKL15 (*lin15(+)*) at 30 ng µl⁻¹ and 1 kb ladder (New England Biolabs) up to 100 ng µl⁻¹. The pBP34 fosmid was injected at 1 ng µl⁻¹ with pBP33 (*P_unc-47::snb-1–GFP*) at 30 ng µl⁻¹, pEKL15 (*lin15(+)*) at 30 ng µl⁻¹ and 1 kb ladder up to 100 ng µl⁻¹.

For tissue-specific expression of the isoforms of *madd-4* or *madd-4–GFP* in cholinergic motoneurons, GABAergic motoneurons or muscle cells, pBP8, pBP11, pBP23, pMP1, pMP9 or pBP31 was injected at 10 ng µl⁻¹ with pEKL15 (*lin15(+)*) at 30 ng µl⁻¹ and 1 kb ladder up to 100 ng µl⁻¹. pBP23 and pMP1 plasmids were also co-injected at 10 ng µl⁻¹ each with pEKL15 (*lin15(+)*) at 30 ng µl⁻¹ and 1 kb ladder up to 100 ng µl⁻¹. pBP70 was injected at 50 ng µl⁻¹ along with pRF4 at 50 ng µl⁻¹.

For labelling of muscle cells, animals were injected with a DNA mixture containing *P_him-4::mb-yfp*[19] at 10 ng µl⁻¹, pEKL15 (*lin15(+)*) at 30 ng µl⁻¹ and 1 kb ladder up to 100 ng µl⁻¹.

For the generation of *madd-4(kr270)* by *Mos*DEL[16,34], the pBP6 repair template vector was injected at 50 ng µl⁻¹ with pJL43 (*Mos* transposase under a germline-specific promoter) at 50 ng µl⁻¹, *P_rab-3::GFP* (a pan-neuronal marker) at 15 ng µl⁻¹ and *P_myo-2::GFP* (a pharynx-specific marker) at 5 ng µl⁻¹.

For the generation of *krSi4* by *Mos*SCI[16,35], the pBP27 vector was injected at 50 ng µl⁻¹ with pJL43 at 50 ng µl⁻¹, *P_rab-3::GFP* at 15 ng µl⁻¹ and *P_myo-2::GFP* at 5 ng µl⁻¹. For the generation of *krSi2* by *Mos*SCI, the pHT22.1 vector was injected at 60 ng µl⁻¹ with pJL43 at 50 ng µl⁻¹, *P_rab-3::GFP* at 15 ng µl⁻¹ and *P_myo-2::GFP* at 5 ng µl⁻¹.

Plasmid construction. We generated fosmids for tagging different isoforms of *madd-4* with GFP, using the fosmid recombineering technique[25]. Specifically, we first added a GFP tag at the C terminus of *madd-4*, right after the last residue, using the WRM0626CA02 fosmid, which contained 37 kb of genomic sequences at the *madd-4* locus, as a template. The recombination fragment was generated by amplification of GFP from pBP32 (a vector derived from pBALU1 (ref. 25) by introducing the S65T mutation in the GFP sequence) flanked by 200-bp *madd-4* homology sequences on each side. In a second step, we recombineered the pBP34 fosmid to delete the first exon of *madd-4A* and *madd-4C* and to make a *madd-4B–GFP* reporter (pBP36) or to remove the first exon of *madd-4B* and to generate a fosmid that drives *madd-4A–GFP* and *madd-4C–GFP* expression (pBP35). To achieve these deletions, the recombination fragments were composed of the GalK cassette amplified from pBALU8* (ref. 25) and 100 bp of homology sequences to the pBP34 fosmid.

The following vectors were generated by Gateway cloning (Invitrogen). They bear the cDNA of *madd-4A* and *madd-4B*, which were obtained by RT–PCR on mixed-stage nematodes RNA extracts. The GFP tags were inserted at the C-terminal end. pBP8: *P_unc-17::madd-4C*; pBP11: *P_unc-17::madd-4A*; pBP23: *P_unc-17::madd-4A-GFP*; pMP1: *P_unc-17::madd-4B-GFP*; pMP9: *P_unc-47::madd-4A-GFP*; pBP31: *P_myo-3::madd-4A-GFP*; pBP70: *P_egl-46::madd-4A-GFP*; pBP33: *P_unc-47::snb-1-3Xmyc*. This reporter for GABA motoneuron presynapses was generated by replacing the GFP sequence of pJL35 (*P_unc-47::snb-1/vamp–GFP*) by a 3 × Myc tag and further cloned using a Gateway cloning strategy (Invitrogen).

In these vectors, the promoter fragments were given by pRH81 (*unc-17* promoter, a gift from E. Jorgensen), pEGB06 (*unc-47* promoter, a gift from E. Jorgensen) and pBP15 (*myo-3* promoter, derived from pmyo-3[4-1], a gift from M. Hammarlund). The *unc-54* 3′ UTR ENTRY vector was a gift from M. Hammarlund. A 3.1-kb promoter region of *egl-46* was amplified from genomic DNA[37].

For expression in HEK cells, *madd-4A* and *madd-4B* cDNAs were tagged at the C-terminal end with 3 × Myc or GFP tags and the fusion proteins were driven by the pCMV promoter in a pCDNA3.1 backbone (Invitrogen). The plasmids were as follows: pBP38: *P_cmv::madd-4A-GFP*; pBP39: *P_cmv::madd-4A-3xmyc*; pBP40: *P_cmv::madd-4B-GFP*; pBP41: *P_cmv::madd-4B-3xmyc*. All plasmid sequences are available from the authors on request.

Electrophysiology. Electrophysiological methods were as described previously[38,39]. L-AChR-dependent evoked currents were recorded in the presence of dihydro-β-erythroidine (5 µM), which selectively inhibits N-AChRs. The experimental sample size was estimated by using values from previous wild-type electrophysiology data, using a 90% confidence level of detecting a less than 0.05 significant difference in amplitudes. Mann–Whitney tests were performed to compare the different genotypes.

Antibody injection. Detection of L-AChR at the surface of muscle cells by antibody injection was performed as described[21]. Monoclonal anti-c-Myc-Cy3 antibodies (C6594; Sigma-Aldrich) and polyclonal anti-GFP-Alexa Fluor 555 antibodies (A-31851; Invitrogen) were used.

Immunofluorescence staining. Immunohistochemistry was performed as described[13]. In brief, *C. elegans* were freeze-cracked and processed through acetone/methanol fixation except for LEV-9-T7, for which paraformaldehyde fixation was used. A mix of primary antibodies or of secondary antibodies were incubated overnight at 4 °C. Primary antibodies were used at the following dilutions: anti-UNC-38, 1:500 (ref. 13); anti-UNC-49, 1:500 (ref. 40); anti-VAChT/UNC-17, 1:1,000 (ref. 12); monoclonal anti-GFP (JL-8; Clontech), 1:500; polyclonal rabbit anti-GFP (Molecular Probes A11122; Invitrogen), 1:500; monoclonal 9E10 anti-Myc (11667149001; Roche), 1:500. Secondary antibodies included Cy3-labelled goat anti-rabbit immunoglobulin G (IgG) used at 1:1,000 (Jackson ImmunoResearch Laboratories A10520; Invitrogen), Cy3-labelled goat anti-mouse IgG at 1:1,000 (Jackson ImmunoResearch Laboratories A10521; Invitrogen), Alexa 488–labelled goat anti-rabbit IgG (Molecular Probes A-11008, Invitrogen) at 1:500, Alexa 488–labelled goat anti-mouse IgG at 1:500 (Molecular Probes A-11001; Invitrogen), Dylight 649-conjugated goat anti-mouse IgG at 1:500 (115-495-146; Jackson ImmunoResearch Laboratories) and Dylight 649-conjugated goat anti-rabbit IgG at 1:500 (111-495-144; Jackson ImmunoResearch Laboratories). Immunofluorescence staining experiments were repeated independently at least three times.

Microscopy. Animals were mounted on 2% agarose pads, anaesthetized with 5 µl of M9 buffer containing 100 mM sodium azide and examined with either a Leica 5000B microscope equipped with a spinning disk CSU10 (Yokogawa) and a Coolsnap HQ2 camera, or a Nikon Eclipse Ti equipped with a spinning disk CSUX1-A1 (Yokogawa) and an Evolve EMCCD camera. Image reconstruction and merges were obtained with Image J. All images shown are projections of Z-stacks 2–20 µm deep and, with the exception of Fig. 3 and Extended Data Figs 2g–j and 5a–f, the background was subtracted with the Subtract Background tool of Image J (rolling ball radius 50 pixels). To remove the autofluorescence of intestinal granules in L1 larvae, a projection of Z-stack images acquired in the green channel was subtracted from a projection of Z-stack images of L-AChR (*unc-29(kr208::tagRFP)*), acquired

in the red channel with the Image Calculator tool in Image J (Fig. 3a, b). For pictures of MADD-4L–GFP and MADD-4B–GFP in L1 larvae (*krEx1067* and *krEx1069*; Fig. 3i, j), the blue channel was used in a similar way to minimize the autofluorescence of intestinal granules.

Image quantification. Quantification of synaptic L-AChR at the dorsal cord in the *unc-29(kr208::tagRFP)* knock-in background was achieved by following a previously published protocol, which was initially described for *in vivo* labelling with anti-Myc–Cy3 (Fig. 4a and Extended Data Fig. 8a)[21]. Fluorescence levels were calculated for the anteriormost 30-μm portion of the dorsal cord. Independent transgenic lines were analysed for each genetic background, on different days (two and three independent lines for Fig. 4a and Extended Data Fig. 8a, respectively). For each line, the normality of distribution was controlled by a Kolmogorov–Smirnov test. The lack of significant difference between data obtained from the independent transgenic lines was tested by two-tailed Student's *t*-tests before pooling data together. Data are presented as a percentage of the mean fluorescence relative to that of control lines acquired on the same days. Eventually, two-tailed Student's *t*-tests were performed to compare the data between different genotypes.

To evaluate the level of co-localization between L-AChR and GABA$_A$R, immunostaining using anti-UNC-49 antibody was performed in the *unc-29(kr208::tagRFP)* knock-in background. Anti-UNC-49 antibodies were revealed with a Dylight 649-conjugated secondary antibody. The fluorescence of tagRFP persisted on immunostaining treatment. In a first analysis, we generated one-dimensional projections of *Z*-stacks corresponding to a region of approximately 150 μm of the anterior dorsal cord by using Image J. The background was removed using the Subtract Background plugin of Image J (rolling ball radius 50 pixels). The fluorescence intensity along the cord was evaluated with the Plot Profile plugin. For each channel, the values were normalized to the value of maximal intensity. The corresponding plots are presented in Fig. 5q, v and Extended Data Fig. 8f, k. The distribution of fluorescence intensity in each field was compared by using the mean of the squares of Pearson's coefficient's (r^2). Analysis of variance (ANOVA) tests were performed to analyse data obtained in the different genetic backgrounds. The plot profiles shown in Fig. 2d, h, l, p and Extended Data Fig. 6h were generated with the same tools.

In the analysis presented in Extended Data Fig. 8l–n, the co-localization of synapses containing GABA$_A$R and L-AChR was examined by identifying the peak of fluorescence in one-dimensional profiles of fluorescence in each colour. To avoid detecting the same clusters twice, the peaks used were the local maxima in a moving window 2.5 μm wide. Only large peaks (that is, with an intensity 1 s.d. above the average fluorescence) were considered. The density of L-AChR clusters around GABA$_A$R clusters was then obtained by constructing the histogram of the distances between all pairs of synapses of different types. For statistical comparison, the amplitude of the peak density above the baseline density was compared across genotypes by ANOVA followed by a pairwise Student's test (with Holm adjustments to the *P* values for repeated comparisons). The statistical analysis was performed blind to the genotype.
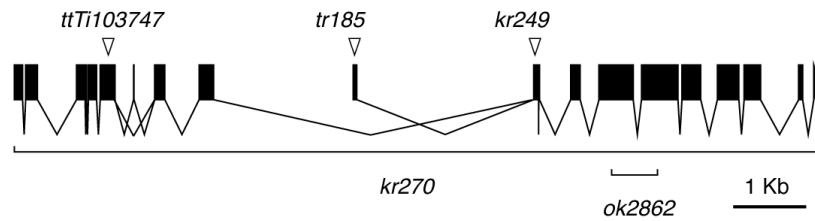
Electron microscopy. High-pressure freezing, freeze substitution, electron microscopy section preparations and electron tomography were performed as described previously[41].

Transfection and Immunoprecipitation. The HEK293 cells were maintained in DMEM medium (Invitrogen) with 10% inactive bovine fetal serum (Invitrogen).
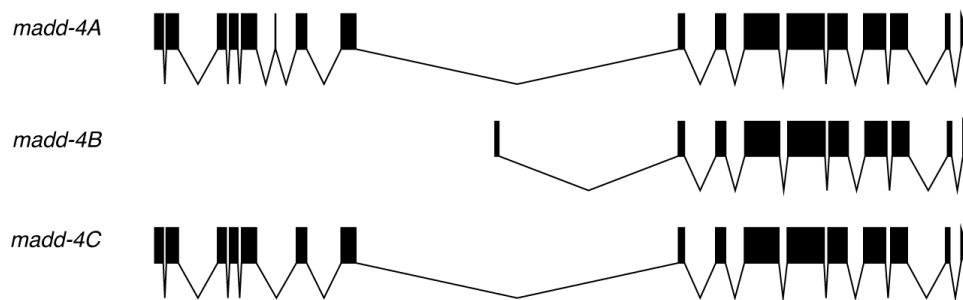
The cells were transiently transfected with indicated constructions (pBP38, pBP39, pBP40 and pBP41) by using jetPRIME transfection reagents (Polyplus). The cell culture supernatants were harvested at 3 days after transfection and used to perform immunoprecipitation after centrifugation to remove the cell debris (13,000 r.p.m. for 10 min at 4 °C). The cells were lysed with ice-cold immunoprecipitation buffer (IPB; 50 mM HEPES pH 7.7, 100 mM NaCl, 50 mM KCl, 2 mM MgCl$_2$, 1 mM EDTA, 1 mg ml^{-1} pepstatin, 1 mg ml^{-1} leupeptin, 2 mg ml^{-1} aprotinin, 1 mM phenylmethylsulphonyl fluoride, one tablet of protease inhibitor cocktail per 10 ml of buffer) containing 1% Triton X-100 on ice for 60 min, then centrifuged at 13,000 r.p.m. for 10 min at 4 °C. Both the cell lysates and cell culture supernatants were pooled together and the mixture was diluted with IPB to a final concentration of 0.2% (w/v) Triton X-100. Further procedure followed a published protocol[21]. In brief, the prewashed 1:1 slurry of 50 μl of anti-GFP-Trap-A beads was added and incubated overnight at 4 °C with gentle rotation. The beads were collected by centrifugation at 1,000*g* for 3 min at 4 °C and washed once with IPB containing 0.2% Triton X-100, and twice with washing buffer (50 mM HEPES pH 7.7, 50 mM NaCl) containing 0.2% Triton X-100. The immunoprecipitated proteins were eluted in Laemmli buffer with 2-mercaptoethanol, and analysed by western blotting after being boiled for 10 min at 95 °C. The primary anti-Myc polyclonal antibody (Gene Tex) was used at a 1:10,000 dilution, and horseradish peroxidase-conjugated goat anti-rabbit (K4002; DAKO) was used as secondary antibody at 1:50 dilution. The images of blotting membranes were captured by Bio-Rad GelDoc XR+ with Image Lab software.

31. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77,** 71–94 (1974).
32. Zhen, M. & Jin, Y. The liprin protein SYD-2 regulates the differentiation of presynaptic termini in *C. elegans*. *Nature* **401,** 371–375 (1999).
33. Sancar, F. *et al.* The dystrophin-associated protein complex maintains muscle excitability by regulating Ca^{2+}-dependent K$^+$ (BK) channel localization. *J. Biol. Chem.* **286,** 33501–33510 (2011).
34. Frokjaer-Jensen, C. *et al.* Targeted gene deletions in *C. elegans* using transposon excision. *Nature Methods* **7,** 451–453 (2010).
35. Frokjaer-Jensen, C. *et al.* Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nature Genet.* **40,** 1375–1383 (2008).
36. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10,** 3959–3970 (1991).
37. Chatzigeorgiou, M. *et al.* Specific roles for DEG/ENaC and TRP channels in touch and thermosensation in *C. elegans* nociceptors. *Nature Neurosci.* **13,** 861–868 (2010).
38. Richmond, J. E., Davis, W. S. & Jorgensen, E. M. UNC-13 is required for synaptic vesicle fusion in *C. elegans*. *Nature Neurosci.* **2,** 959–964 (1999).
39. Richmond, J. E. in *WormBook* (ed. The *C. elegans* Research Community) http://dx.doi.org/10.1895/wormbook.1.112.1 (October 6, 2006).
40. Gally, C. & Bessereau, J. L. GABA is dispensable for the formation of junctional GABA receptor clusters in *Caenorhabditis elegans*. *J. Neurosci.* **23,** 2591–2599 (2003).
41. Stigloher, C., Zhan, H., Zhen, M., Richmond, J. & Bessereau, J. L. The presynaptic dense projection of the *Caenorhabditis elegans* cholinergic neuromuscular junction localizes synaptic vesicles at the active zone through SYD-2/liprin and UNC-10/RIM-dependent interactions. *J. Neurosci.* **31,** 4388–4396 (2011).

a - Locus

ttTi103747 *tr185* *kr249*

kr270

ok2862

1 Kb

b - Transcripts

madd-4A

madd-4B

madd-4C

c - Fosmids

madd-4-GFP

madd-4L-GFP

madd-4B-GFP

d MADD-4A / MADD-4C
(1045 / 1043 amino acids)

MADD-4B
(711 amino acids)

Signal peptide

ADAMTS Cysteine-rich module

ADAMTS spacer module

Thrombospondin type 1 repeat

Ig C2 type domain

PLAC domain (Protease and LACunin)
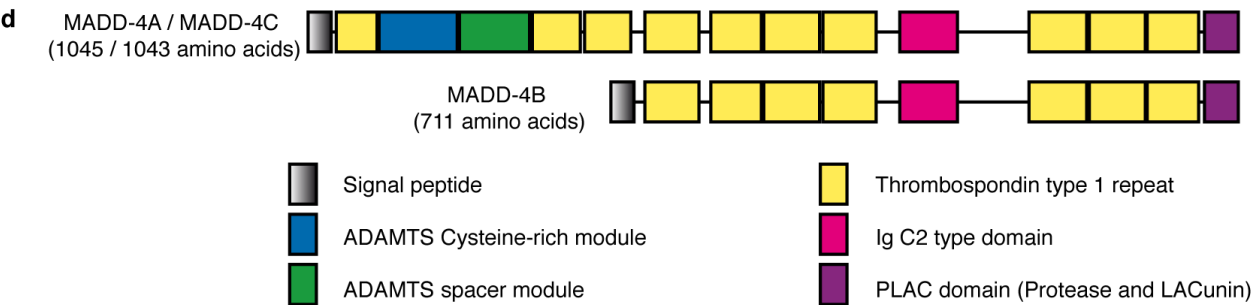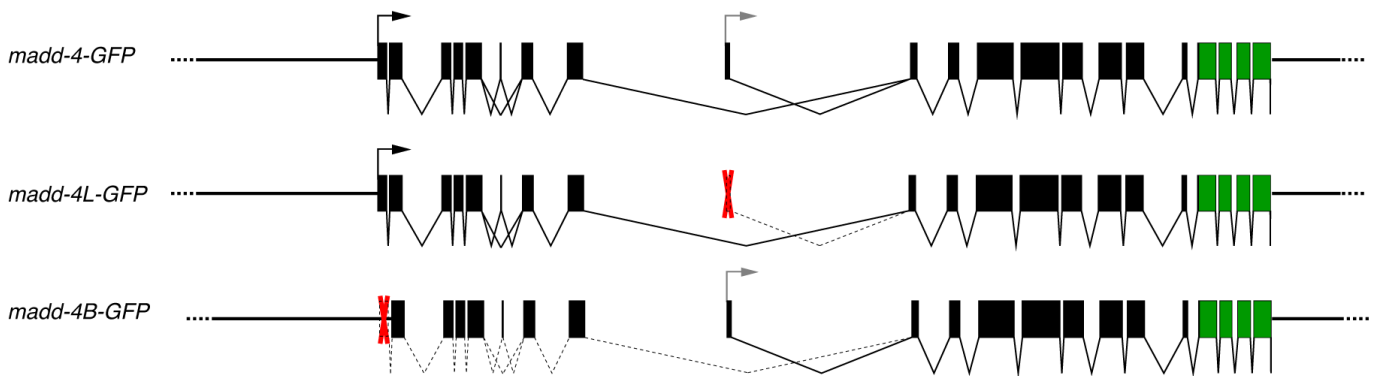
e
```
TSP1-3   CGRWEAGKWSKCTAS-CGQ---GVRRRHVACVGG------------SDCD--EGGRP-RQETTCYAGIPCS
TSP1-4   --RLVAGEWSTCSST-CGT---GVMSRTVECVAVNPISSAPIKLPMSECQ--DQEQP-KLFESCEVRS-CP
TSP1-7   --TWVTAQWTECSRS-CDS---GERRQVWCEIRDSRGKTQ-RRPDVECD--ANTKP-QTVEVCSFGS-CS
TSP1-9   ---WEASQWSECSSQRCVSSMLAQKRRNVTCRFTNGTS-----VDIQHCD--ITNRP-ATTMDCPNQN-CK
TSP1-5   ---WRYGDWTQCSAS-CLG---GKQKAALKCIQVSTGKS----VQWSQCD--ARRRPPEKSRPCNQHP-CP
TSP1-8   --RWDIGHWSECRQKTCHVA--GYQARGISCKVTFHGEIRN--VDNSICESLASVRP-PETRPCHRED-CP
          ..*: *      *        .      : *          *:      :*        *        *
```

Extended Data Figure 1 | Structure of the *madd-4* locus and of the fosmid reporters and domain composition of long and short MADD-4 isoforms.
a, *madd-4* genomic organization (black boxes represent exons), and position of different mutants used in this study (open arrowheads represent insertion and point mutations; brackets represent deletions). One allele disrupts the long isoform specifically: *madd-4(ttTi103747)* has a *Mos*1 insertion in the fifth exon of *madd-4A* and *madd-4C* (*Mos*1 insertion at position 601 of *madd-4A* cDNA, based on Wormbase release WS240). One allele disrupts the short isoform *madd-4(tr185)* specifically: M1I (ref. 5). The other alleles affect all isoforms: *madd-4(kr249)*, W374stop (A instead of G at position 1121 of *madd-4A* cDNA, this allele was isolated based on a L-AChR distribution defect following mutagenesis); *madd-4(ok2862)*, 631-bp deletion at position 1438 of *madd-4A* cDNA leading to an insertion of 16 amino acids and a premature stop[5]; and *madd-4(kr270)*, 11,042-bp full deletion of the locus (from the ATG of *madd-4A* to the stop codon) engineered by homologous recombination. Because

madd-4(kr249) shows the same L-AChR and GABA$_A$R clustering defects as *madd-4(kr270)*, it is probably a null allele. Scale bar, 1 kb. **b**, The three mRNAs transcribed from the *madd-4* locus. In this paper, MADD-4L refers to both MADD-4A and MADD-4C, in instances where they cannot be distinguished. **c**, Structure of the *madd-4* locus in the three fosmids obtained by recombineering of the WRM0626CA02 fosmid (see Methods). Arrows indicate the beginning of the open reading frame of *madd-4L* (black) and *madd-4B* (grey). Green exons correspond to the sequence of the GFP. Red crosses indicate removal of exons. **d**, The different domains of MADD-4A, MADD-4C and MADD-4B proteins. The two long isoforms (MADD-4A and MADD-4C, globally referred to as MADD-4L) have the same domain composition. **e**, Sequence alignment of six TSP-1 repeats of MADD-4L shows six conserved cysteine residues, characteristic of these domains. The TSP-1 no. 9 was not described previously[5]. The sequence alignment was generated with Clustal W (EBI).
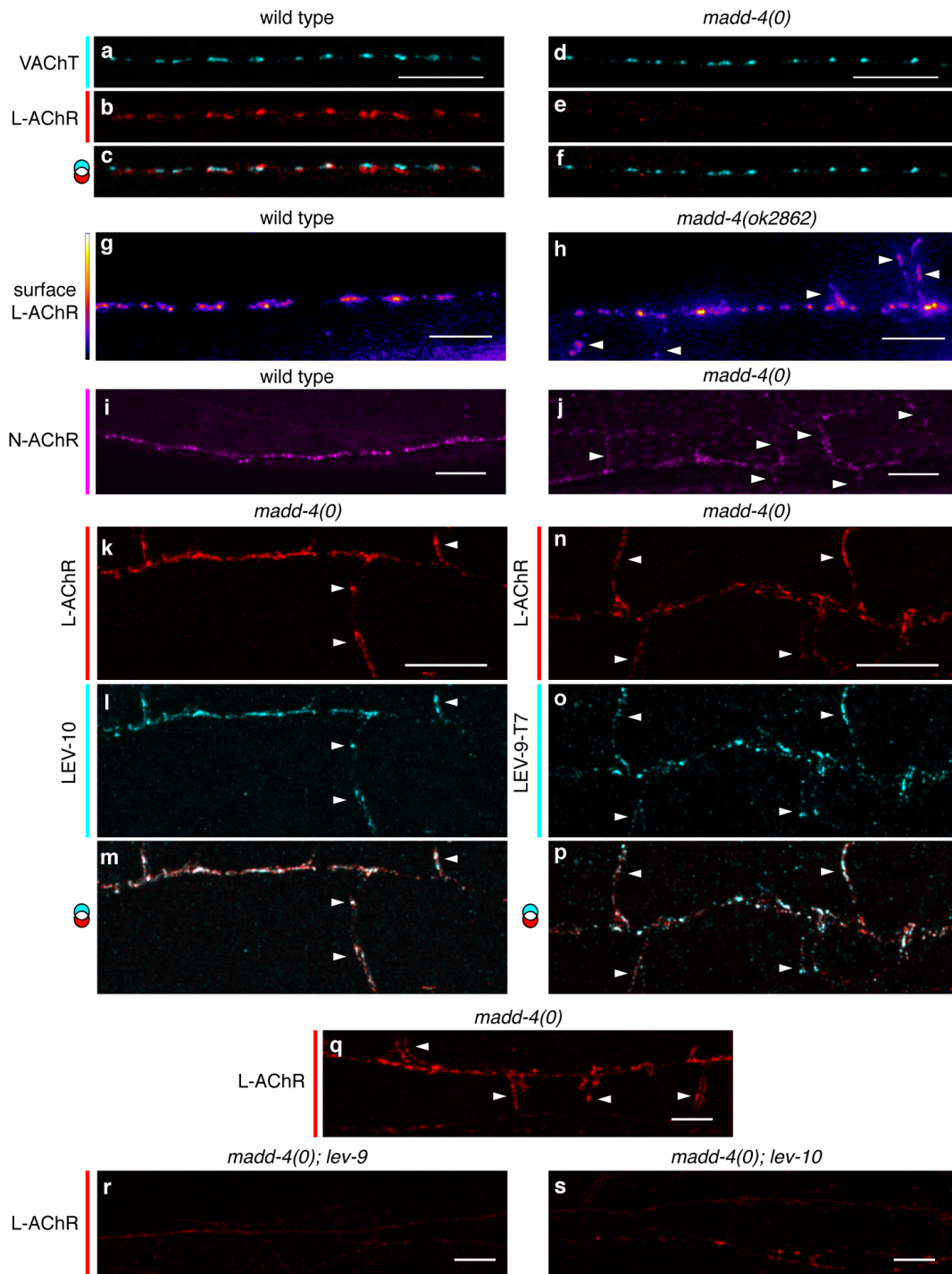
Extended Data Figure 2 | Both L- and N-AChRs are redistributed at extrasynaptic areas at the surface of muscle cells, and L-AChRs remain associated with scaffolding proteins in *madd-4* mutant.
a–f, SAB head motoneurons differentiate wild-type cholinergic boutons (immunofluorescence, anti-VAChT/UNC-17 antibody) in a *madd-4(kr270)* null mutant (**a**, **d**) but synaptic L-AChRs (immunofluorescence, anti-UNC-38) are not detectable (**b**, **e**). **g**, **h**, Transgenic animals expressing a Myc-tagged UNC-38 L-AChR subunit were injected with fluorescent anti-Myc antibodies into the body cavity to reveal L-AChRs at the muscle cell surface, in the wild type (**g**) and in *madd-4(ok2862)* mutant background (**h**). **i**, **j**, N-AChRs

(ACR-16–GFP) are redistributed to muscle arms in a *madd-4(kr270)* null mutant (**j**), in contrast with the wild type (**i**). **k–p**, At the dorsal cord of a *madd-4(kr249)* mutant, synaptic and non-synaptic L-AChR clusters (*unc-29–RFP* in red) (**k**, **n**) co-localize with LEV-10 detected by immunofluorescence (cyan) (**l**) and with LEV-9 (**o**) stained with anti-T7 antibody in *unc-29–RFP madd-4(kr249); lev-9-T7* knock-in animals. **q–s**, L-AChR clusters (*unc-29–RFP*) remaining in *madd-4(kr249)* are lost in *lev-9* and *lev-10* mutant backgrounds. Arrowheads point to extrasynaptic AChR clusters. **c**, **f**, **m**, **p**, Channel overlays along with a colour mixing guide. Scale bars, 10 μm.
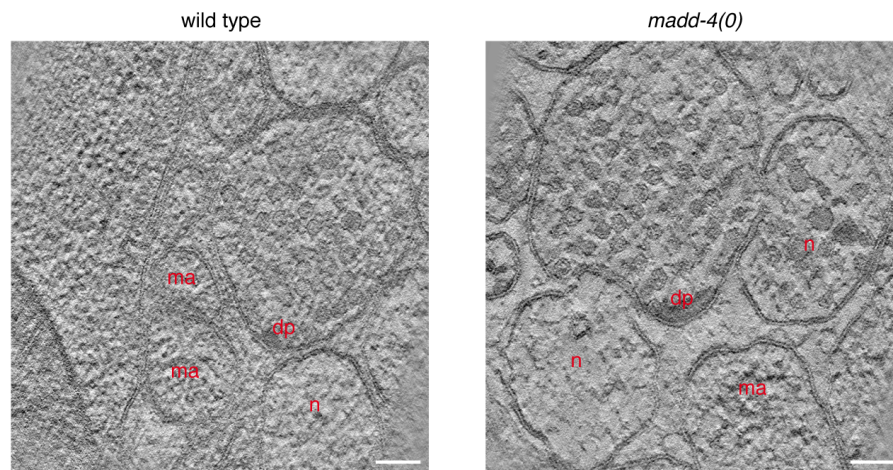
wild type

madd-4(0)

Extended Data Figure 3 | Synapse morphology is normal at the ultrastructural level in the *madd-4* mutant. Electron microscopy tomogram sections of cholinergic NMJs in the dorsal nerve cords of wild-type and *madd-4(kr249)* null mutant. ma, muscle arm; dp, dense projection; n, neuron. Scale bars, 100 nm.
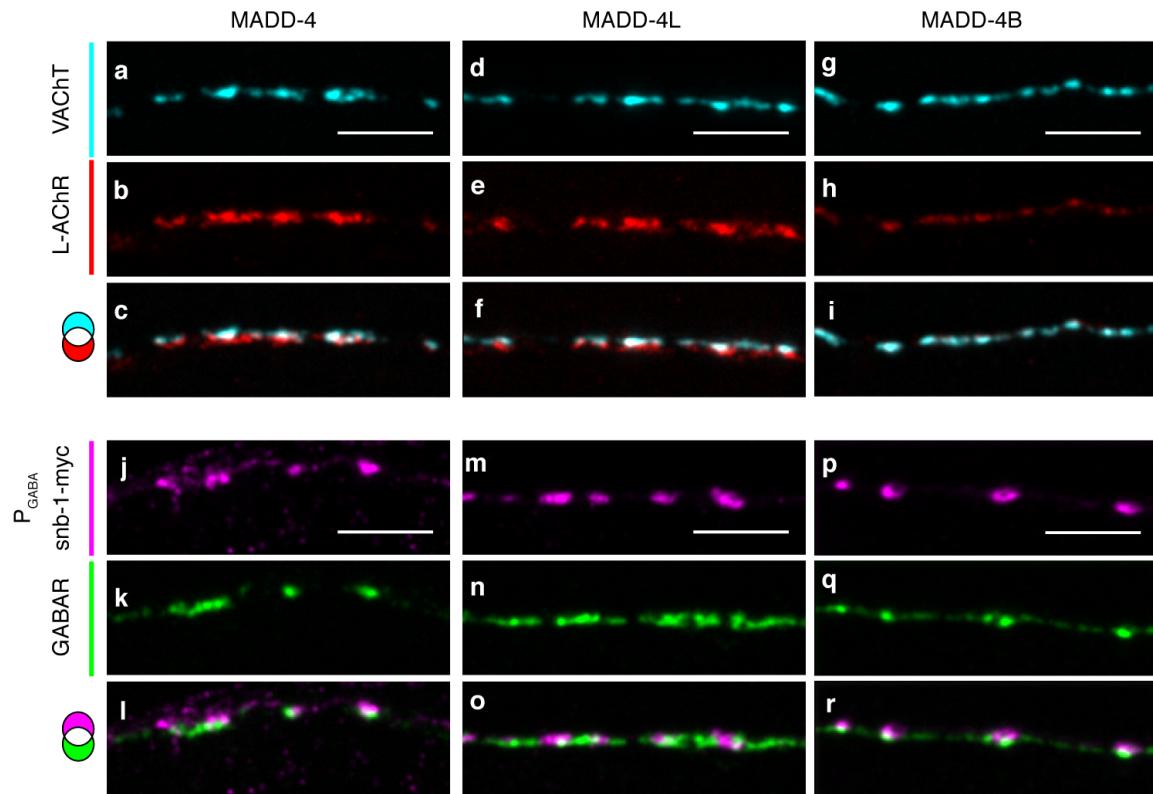
Extended Data Figure 4 | Analysis of L-AChR and GABA$_A$R localization in transgenic lines expressing MADD-4L or MADD-4B. Immunofluorescence analysis of transgenic *madd-4(kr249)* null mutants containing the recombineered fosmids for revealing MADD-4–GFP (**a–c, j–l**), MADD-4L–GFP (**d–f, m–o**) or MADD-4B–GFP (**g–i, p–r**) (see Extended data Fig. 1c). L-AChRs (anti-UNC-38 antibody (**b, e, h**)) are localized opposite cholinergic boutons (anti-VAChT/UNC-17 antibody (**a, d, g**)) at a wild-type level when MADD-4–GFP or MADD-4L–GFP is expressed, and at a lower level when MADD-4B–GFP is expressed. Immunofluorescence staining of UNC-49 shows a rescue of GABA$_A$Rs for MADD-4–GFP (**k**) and for MADD-4B–GFP (**q**) but not for MADD-4L–GFP (**n**). GABA boutons are revealed with a P$_{unc-47}$::snb-1-myc marker (**j, m, p**). **c, f, i, l, o, r**, Channel overlays along with a colour mixing guide. Scale bars, 10 μm.
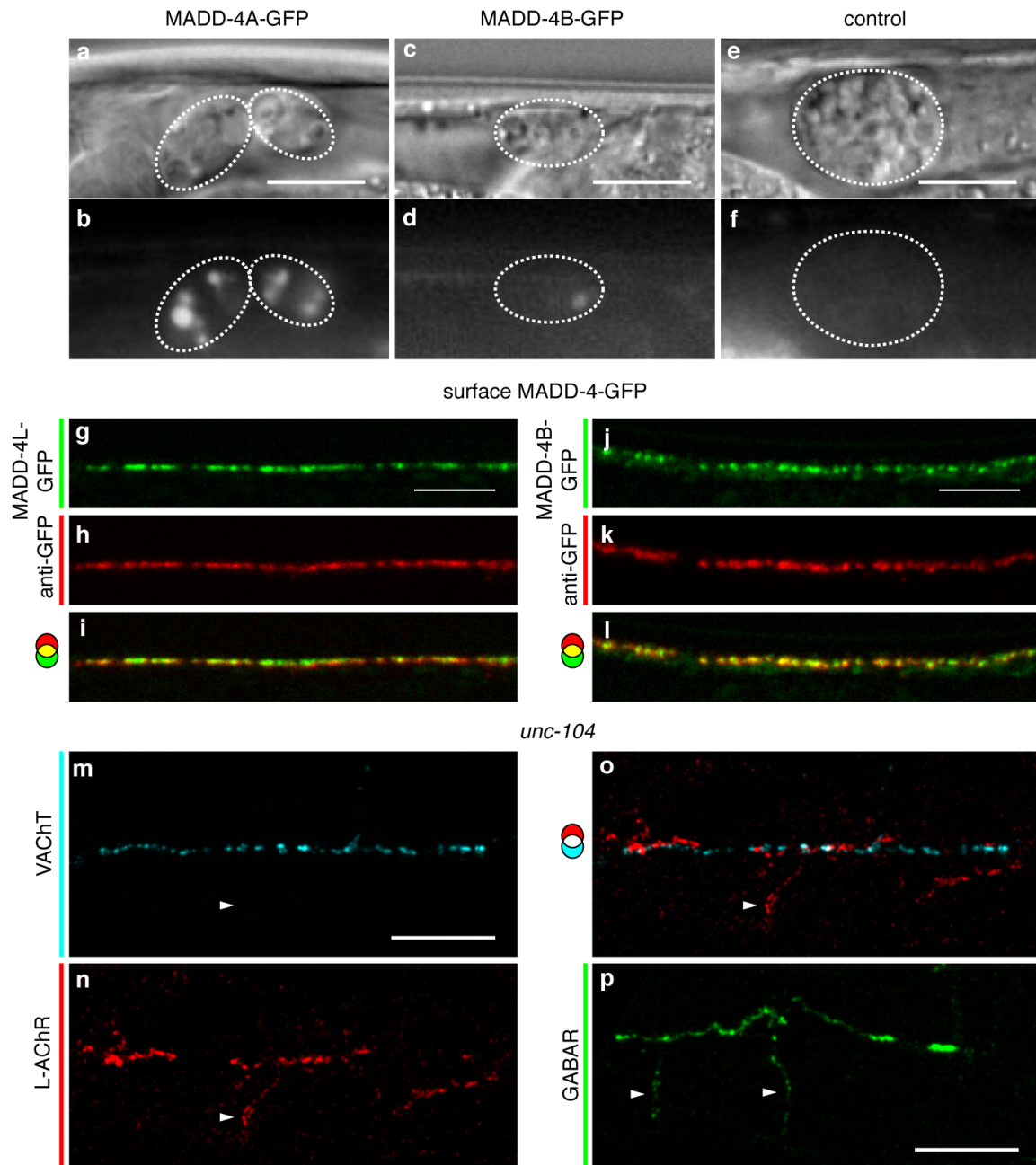
Extended Data Figure 5 | MADD-4 is secreted by motoneurons and transported along axons. a–f, In transgenic strains expressing MADD-4A–GFP (a, b) and MADD-4B–GFP (c, d) from recombineered fosmids, fluorescence is detected in intracellular vesicles of coelomocytes (highlighted with white dotted circles). This indicates that MADD-4–GFP proteins were secreted in the pseudocoelomic cavity and endocytosed by these scavenger cells, which filter the pseudocoelomic fluid and are not fluorescent in the wild type (e, f). g–l, Transgenic animals expressing GFP-tagged MADD-4L (g) or MADD-4B (j) in the madd-4(kr270) mutant background were injected into the body cavity with red-fluorescent anti-GFP antibodies (conjugated with

Alexa 555) (h, k) to reveal MADD-4 isoforms secreted at synapses. m–o, Immunofluorescence staining shows that L-AChRs (anti-UNC-38 antibody) (n) are redistributed at extrasynaptic areas at the dorsal side in the unc-104(e1265) mutant, as in a madd-4 null mutant. Cholinergic boutons are revealed with an anti-VAChT/UNC-17 antibody (m). p, Immunofluorescence staining of UNC-49 in unc-104(e1265) mutants shows a redistribution of GABA$_A$Rs in muscle arms at the dorsal side, which mimics the loss of madd-4. Arrowheads show extrasynaptic clustering. i, l, o, Channel overlays along with a colour mixing guide. Scale bars, 10 μm.
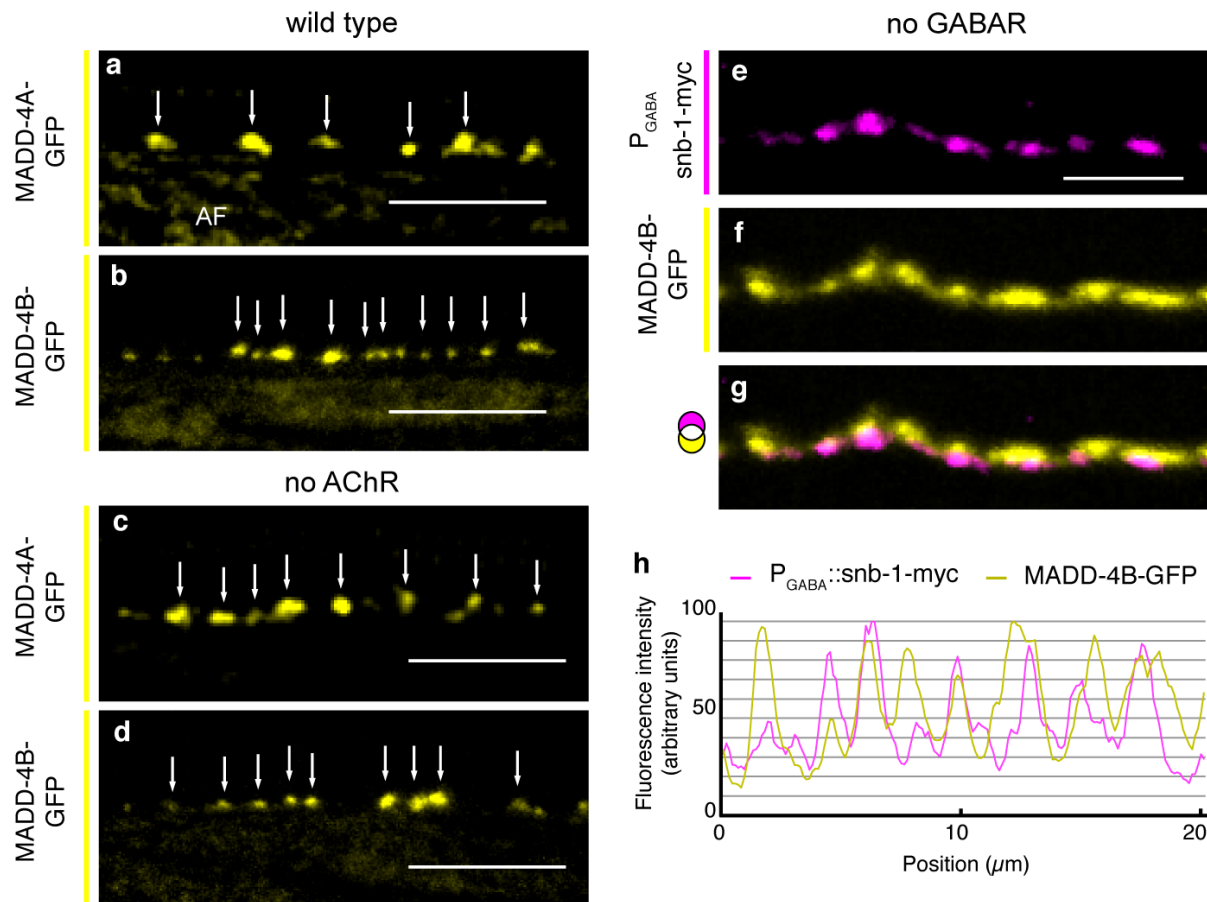
Extended Data Figure 6 | MADD-4 isoforms are retained at synapses independently from cholinergic and GABAergic receptors. a–d, MADD-4A–GFP (*krSi4[P_{madd-4L}::madd-4A–GFP]*) (**a,c**) and MADD-4B–GFP (*krEx1069[madd-4B–GFP fosmid]*) (**b, d**) are unaffected by the absence of L-AChRs and N-AChRs in an *unc-29(x29); acr-16(ok789)* double mutant (**c, d**), when compared with the wild type (**a, b**). Arrows indicate the position of synapses made with SAB motoneurons. AF, non-specific autofluorescence of muscle cells. **e–h**, MADD-4B–GFP (**f**) is juxtaposed to dorsal cord GABA boutons (*P_{unc-47}::snb-1-myc* marker) (**e**) in the absence of GABA_ARs in the *unc-49(e407)* mutant (**e–h**), as in the wild type (Fig. 2m–p). **g**, Channel overlay along with a colour mixing guide. Scale bars, 10 μm (**a–d**) and 5 μm (**e–g**).
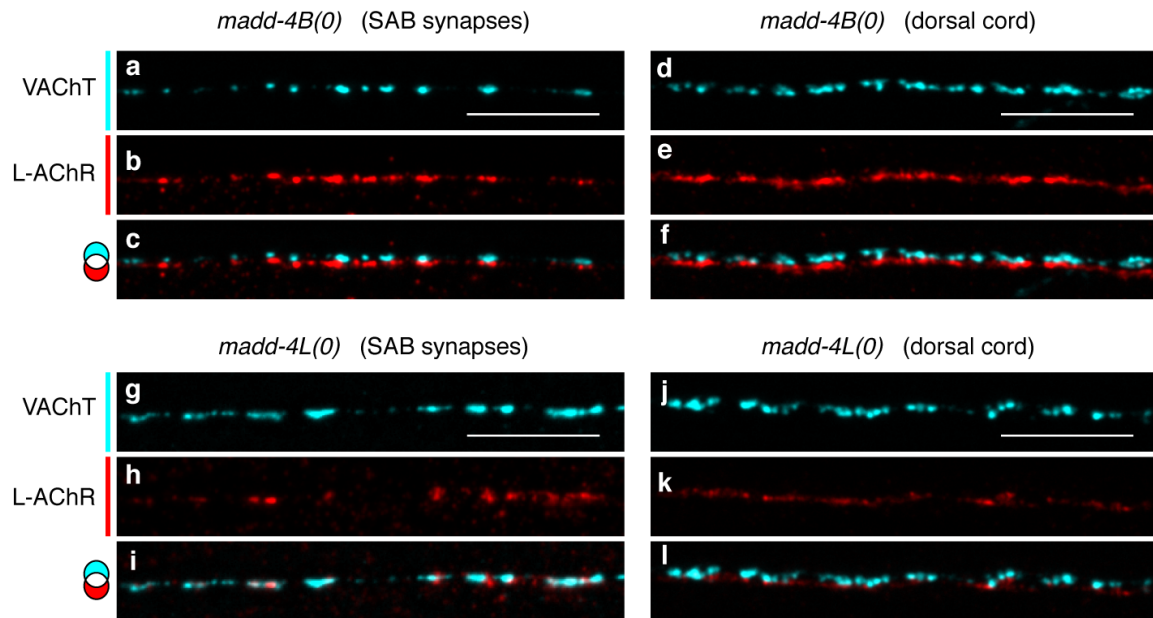
Extended Data Figure 7 | Analysis of L-AChR localization in *madd-4B(0)* and *madd-4L(0)* mutants. Immunofluorescence staining shows that L-AChRs (anti-UNC-38 antibody) (**b, e, h, k**) are clustered opposite acetylcholine release sites (anti-VAChT/UNC-17 antibody) (**a, d, g, j**) at a wild-type level in *madd-4B(0)* mutants (*madd-4(tr185)*) (**a–f**) and at a lower level in *madd-4L(0)* mutants (*madd-4(ttTi103747)*) (**g–l**), both at synapses made with SAB motoneurons (**a–c, g–i**) and at the dorsal nerve cord (**d–f, j–l**). **c, f, i, l,** Channel overlays along with a colour mixing guide. Scale bars, 10 μm.
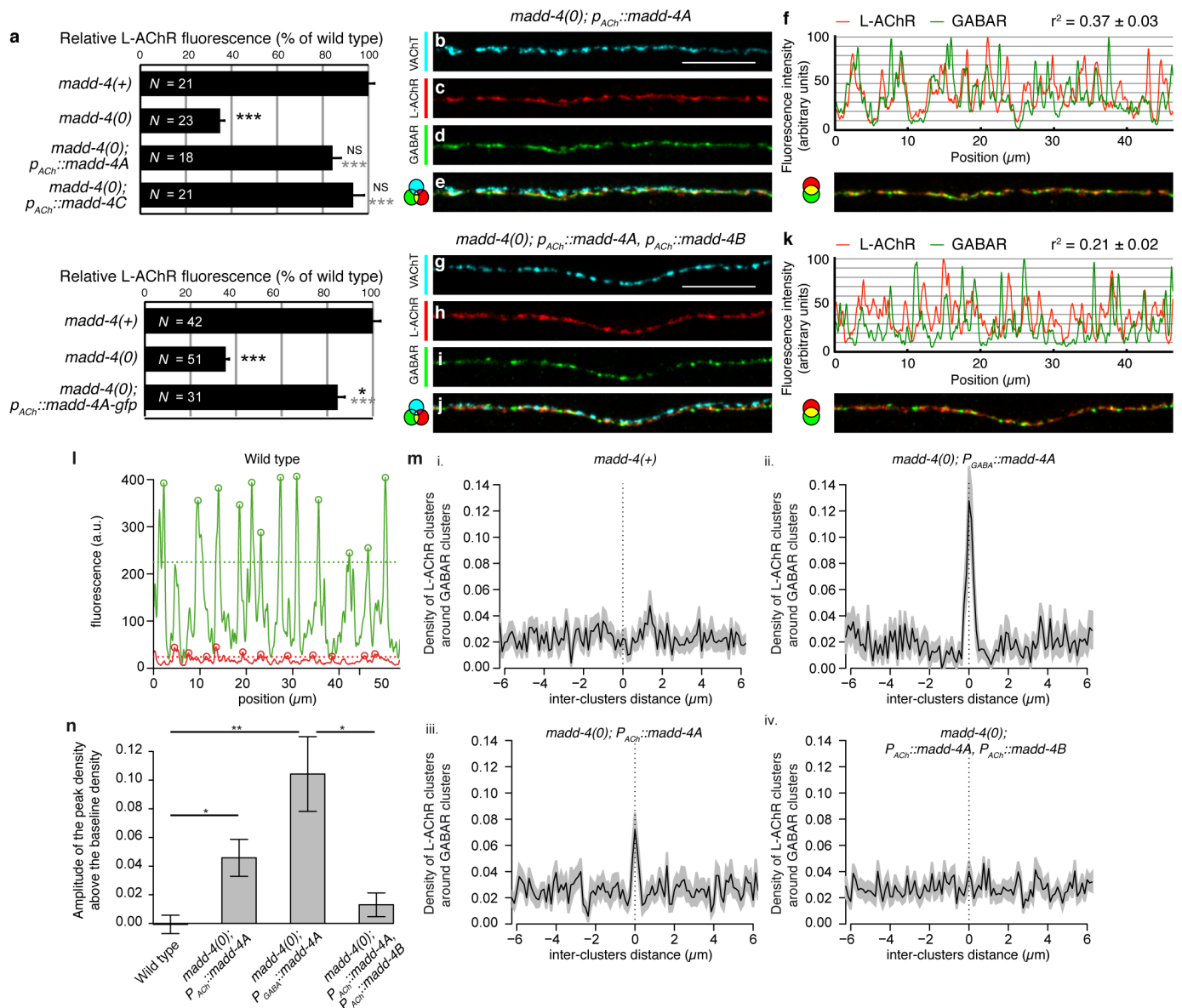
Extended Data Figure 8 | Analysis of L-AChR and GABA_AR distribution in transgenic lines expressing different combinations of MADD-4 isoforms. **a**, Synaptic clustering of L-AChRs (*unc-29*–RFP knock-in) was evaluated in the control *madd-4(+)* and in transgenic *madd-4(kr249)* mutants animals expressing MADD-4A, MADD-4C or MADD-4A–GFP in cholinergic motoneurons. Data are presented as means and s.e.m.; *N* = number of worms; Student's *t*-test compared with the wild type (black) or compared with *madd-4(kr249)* (grey): ***$P \leq 0.001$; *$P \leq 0.01$; NS, not significant. **b–k**, Labelling of cholinergic terminals (anti-UNC-17 immunofluorescence) (**b, g**), L-AChRs (UNC-29–RFP) (**c, h**) and GABA_ARs (anti-UNC-49 immunofluorescence) (**d, i**) in *madd-4(kr249)* null mutants expressing the long isoform MADD-4A [*madd-4(0); P_{ACh}::madd-4A*] or both MADD-4A and MADD-4B [*madd-4(0); P_{ACh}::madd-4A P_{ACh}::madd-4B*] in cholinergic motoneurons. The relative fluorescence intensity of L-AChRs and GABA_ARs along the cord is shown (**f, k**). Co-localization of L-AChRs and GABA_ARs was quantified in 25 and 33 worms, respectively. r^2 values are the means of squares of Pearson's coefficient, r^2 (mean ± s.e.m.). A value close to 0 indicates a lack of co-localization, whereas a value close to 1 indicates complete co-localization because all calculated *r* values were positive. The r^2 of the four genotypes (also presented in Fig. 5q, v) is statistically different by ANOVA. Scale bars, 10 μm. **l–n**, L-AChRs (UNC-29–RFP) and GABA_ARs (anti-UNC-49 immunofluorescence) were detected

together in *madd-4(+)* and in *madd-4(kr249)* mutants expressing the long isoform MADD-4A in GABAergic motoneurons [*madd-4(0); P_{GABA}::madd-4A*], in cholinergic motoneurons [*madd-4(0); P_{ACh}::madd-4A*] or expressing both MADD-4A and MADD-4B in cholinergic motoneurons [*madd-4(0); P_{ACh}::madd-4A P_{ACh}::madd-4B*]. **l**, Example of the fluorescence profile of GABA_ARs (green) and L-AChRs (red) of *madd-4(+)*; receptor clusters are detected as local peaks in the fluorescence profile 1 s.d. above the mean fluorescence (see Methods). The threshold for the green and red profiles is indicated by the dashed lines; the circles indicate the position of the peaks used for the co-localization study. **m**, Average-density histogram of L-AChR clusters around GABA_AR clusters for the four genotypes. In *madd-4(+)*, the flat curve reveals a random distribution of cholinergic NMJs relative to GABAergic NMJs. The co-localization of GABA_ARs and L-AChRs is revealed by a central peak in the histogram. Number of worms used: *N* = 23, *N* = 19, *N* = 25 and *N* = 33, respectively. Bin = 0.125 μm (2 pixels); mean (plain line) ± s.e.m. (grey area). **n**, Quantification of the amplitude of the central peaks in the histogram density of **b** (expressed as a deviation from the average density). The genotypes are statistically different (ANOVA on the factor genotype: $F(3,96) = 10.46$, $P < 10^{-4}$). Post-hoc Student's test: *$P < 0.05$; **$P < 0.01$. **e, f** (lower), **j, k** (lower), Channel overlays along with a colour mixing guide.
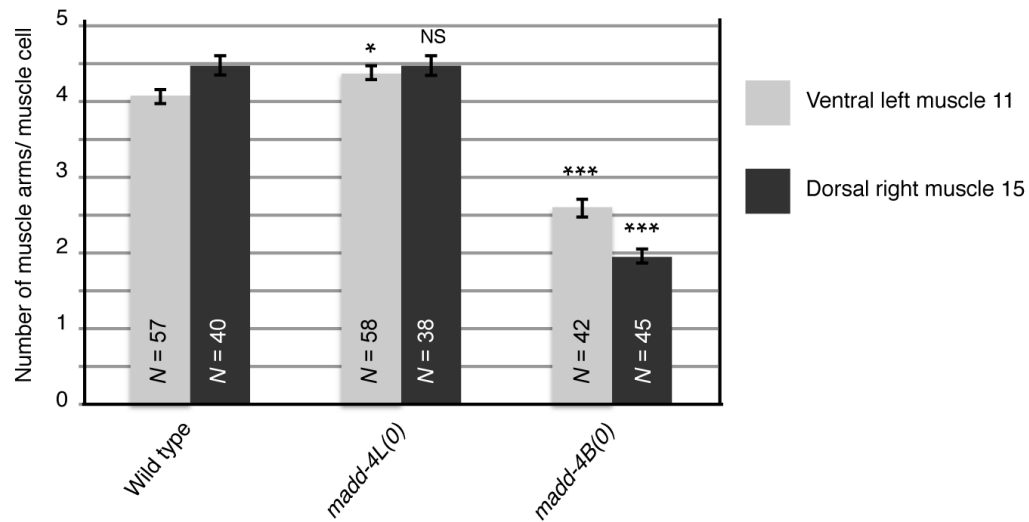
Extended Data Figure 9 | *madd-4B(0)* mutants show muscle arm extension defects but *madd-4L(0)* mutants do not. The average number of muscle arms for the ventral left muscle 11 and the dorsal right muscle 15 is indicated for the wild type and the *madd-4(ttTi103747)* (*madd-4L(0)*) and *madd-4(tr185)* (*madd-4B(0)*) mutants. Distal muscle cells were revealed with a $P_{him-4}::yfp$ muscle arm reporter (*trIs25*). Results are shown as means ± s.e.m.; N = number of worms; Mann–Whitney test: ***$P \le 0.0001$; *$P \le 0.05$; NS, not significant.
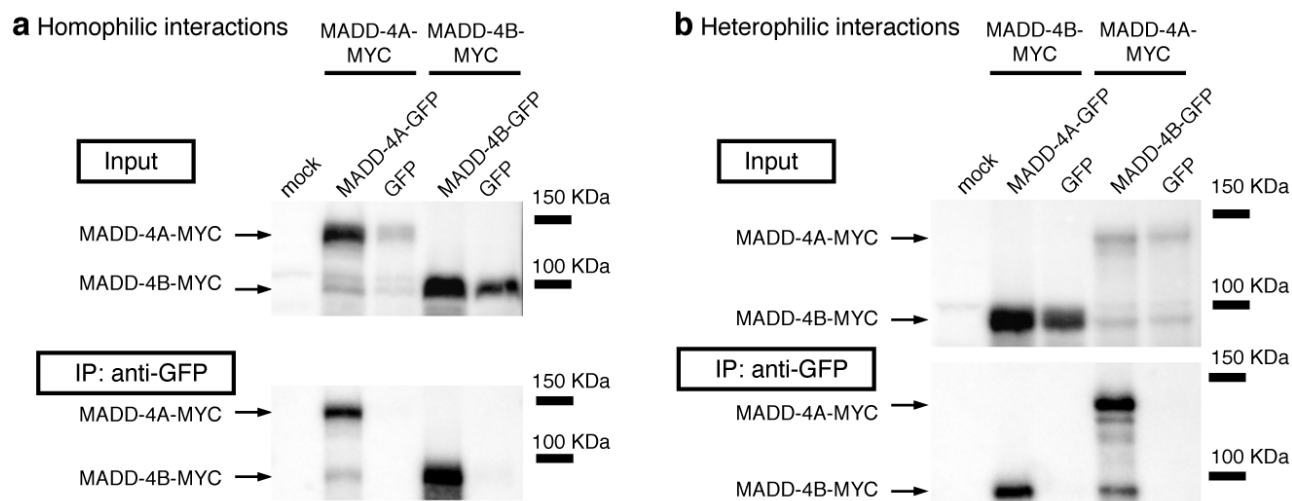
Extended Data Figure 10 | MADD-4 isoforms undergo homophilic and heterophilic interactions. a, Immunoprecipitation of MADD-4A–GFP (lane 2) or MADD-4B–GFP (lane 4) with anti-GFP antibodies was followed by western blot analysis with anti-Myc antibodies to detect MADD-4A–Myc or MADD-4B–Myc. A specific MADD-4A or MADD-4B signal was detected in the immunoprecipitation products of MADD-4A–GFP or MADD-4B–GFP, respectively, but not when GFP alone was co-expressed as a control (lanes 3 and 5). The GFP-tagged and Myc-tagged proteins were co-expressed in HEK cells. Lane 1, mock control. **b**, By using a similar procedure, a specific MADD-4A or MADD-4B signal was detected in the immunoprecipitation products of MADD-4B–GFP or MADD-4A–GFP, respectively. These experiments were repeated independently three times.

LETTER

Modality–specific thalamocortical inputs instruct the identity of postsynaptic L4 neurons

Gabrielle Pouchelon[1], Frédéric Gambino[1]†, Camilla Bellone[1], Ludovic Telley[1], Ilaria Vitali[1], Christian Lüscher[1,2,3], Anthony Holtmaat[1] & Denis Jabaudon[1,2,3]

During development, thalamocortical (TC) input has a critical role in the spatial delineation and patterning of cortical areas[1–6], yet the underlying cellular and molecular mechanisms that drive cortical neuron differentiation are poorly understood. In the primary (S1) and secondary (S2) somatosensory cortex, layer 4 (L4) neurons receive mutually exclusive input originating from two thalamic nuclei[7,8]: the ventrobasalis (VB), which conveys tactile input[9,10], and the posterior nucleus (Po), which conveys modulatory and nociceptive input[11–14]. Recently, we have shown that L4 neuron identity is not fully committed postnatally[15], implying a capacity for TC input to influence differentiation during cortical circuit assembly. Here we investigate whether the cell-type-specific molecular and functional identity of L4 neurons is instructed by the origin of their TC input. Genetic ablation of the VB at birth resulted in an anatomical and functional rewiring of Po projections onto L4 neurons in S1. This induced acquisition of Po input led to a respecification of postsynaptic L4 neurons, which developed functional molecular features of Po-target neurons while repressing VB-target traits. Respecified L4 neurons were able to respond both to touch and to noxious stimuli, in sharp contrast to the normal segregation of these sensory modalities in distinct cortical circuits. These findings reveal a behaviourally relevant TC-input-type-specific control over the molecular and functional differentiation of postsynaptic L4 neurons and cognate intracortical circuits, which instructs the development of modality-specific neuronal and circuit properties during corticogenesis.

Within S1, VB axons target L4 neurons, forming cortical barrels, and Po axons target L5A and L1 neurons; in S2, however, Po axons target L4 neurons (Fig. 1a, b)[7,8,16,17]. Molecular distinctions between L4 neurons in S1 and S2, such as *Rorb* expression[18] (Fig. 1c), may therefore in part be driven by their distinct TC inputs. To investigate TC-type-specific controls over L4 neuron identity, we genetically ablated the VB at birth by generating transgenic *Slc6a4*::Cre/*Rosa26*::stop[flox]*DTA* mice (*vb⁻* mice)[19,20], leading to death of VB neurons between postnatal day (P)0 and P4 and lack of formation of associated S1 whisker-mapped cortical barrels (Fig. 1d–h and Extended Data Figs 1, 2a–c). Remarkably, S1L4 remained richly innervated by TC terminals despite the absence of VB axons, as shown using the pan-TC presynaptic marker VGLUT2 (ref. 18) (Fig. 1i), without evidence for secondary cell death (Extended Data Fig. 2d–g). These terminals did not belong to residual VB axons, as they were still present at P23 (Extended Data Fig. 2b) and did not express the VB-specific markers 5HTT (5-hydroxytryptamine transporter) and GSBS (G-substrate)[21] (Fig. 1j, k). These data indicate that S1L4 neurons still receive TC input in the absence of VB.

To identify the origin of these ectopic TC projections, we retrogradely labelled TC neurons from S1, revealing that Po is the exclusive source of TC input to S1 in *vb⁻* mice (Extended Data Fig. 3a–f). Po$_{vb-}$ neurons were undistinguishable from control Po neurons by microarray comparative gene expression analysis (Extended Data Fig. 3g), demonstrating that bona fide Po neurons are the source of S1L4 TC input in *vb⁻* mice. To directly visualize aberrant Po$_{vb-}$ projections, we next anterogradely
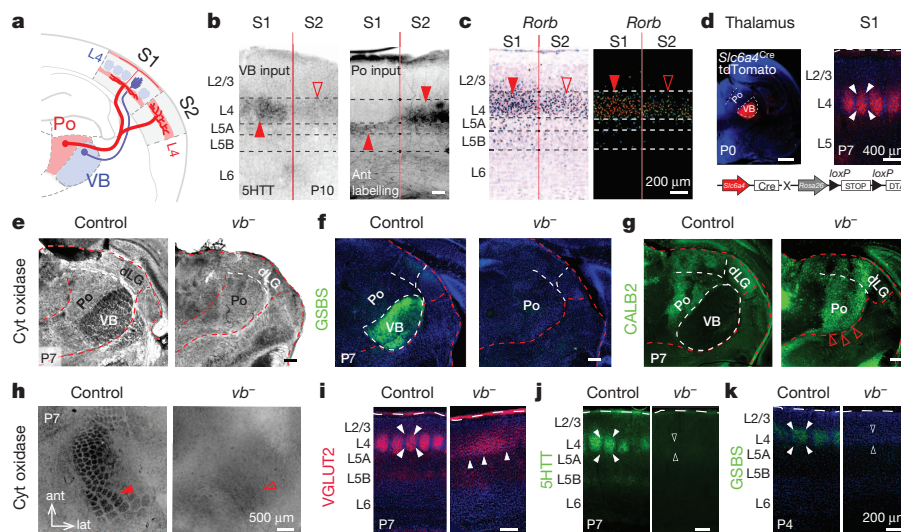
Figure 1 | TC input to S1L4 is preserved in *vb⁻* mice. a, b, VB and Po TC axons have specific projections to S1 and S2. **c,** *Rorb* expression is area-specific in L4 neurons (from Allen Brain Atlas). **d,** tdTomato reporter expression driven by *Slc6a4*^Cre shows VB-specific recombination. **e–h,** *vb⁻* mice lack VB (**e–g**) and lack presynaptic barrels in S1 (**h**). Ant, anterograde; Cyt oxidase, cytochrome oxidase; dLG, dorsolateral geniculate nucleus. **i,** Presynaptic TC terminals are still present in S1$_{vb-}$ cortex using the pan-TC presynaptic marker VGLUT2. **j, k,** VB-specific presynaptic markers are lacking in S1$_{vb-}$.

[1]Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, CH-1211 Geneva, Switzerland. [2]Clinic of Neurology, Department of Clinical Neurosciences, Geneva University Hospital, CH-1211 Geneva, Switzerland. [3]Institute of Genetics & Genomics in Geneva (iGE3), University of Geneva, CH-1211 Geneva, Switzerland. †Present address: Interdisciplinary Institute for NeuroScience, CNRS UMR 5297, 33077 Bordeaux, France.

labelled Po and VB axons in control mice and Po axons in vb^- mice. Control VB axons branched within S1L4 and Po axons were excluded from this layer, whereas conversely, in S2, where VB axons were not found, Po axons branched within L4 (Fig. 2a, b, d). By contrast, in $S1_{vb-}$, Po neurons projected densely to L4 (Fig. 2c, d; $P < 0.05$ for Po control (Po_{Ctrl}) versus Po_{vb-} axonal length in L4, Student's t-test, $n = 3$ control and $n = 4$ vb^- labellings). Po axons formed functional synapses onto $S1L4_{vb-}$ neurons, as demonstrated by short-latency synaptic responses to optogenetic stimulation of these axons, which were absent in control S1L4 barrel neurons (Fig. 2e, f). By contrast, Po-S2L4 connectivity was unchanged in vb^- cortex (Extended Data Fig. 3h, i), and Po rewiring in S1 was not observed following focal VB ablation at P10 (Supplementary Note 1). Taken together, these results indicate that Po axons substitute for VB axons in vb^- mice, providing a new source of presynaptic input to L4 neurons.

We next investigated whether this VB→Po switch in TC input instructs cell-type-specific developmental gene expression programs in postsynaptic L4 neurons. We first characterized the molecular identity of wild-type cortical neurons in L2/3, L4, and L5/6 of S1 and S2 using microarray analysis of microdissected samples[22] at P10 (Fig. 3a, b and Extended Data Fig. 4a, b). The genetic relatedness between these samples was determined using unbiased cluster analysis, revealing three classes of transcriptional programs: L2/3-like, L4-like and L5/6-like. Analysis of $S1L4_{vb-}$ samples revealed an L4-like transcriptional program, demonstrating that TC input origin does not determine the laminar molecular identity of L4 neurons (Fig. 3c).

We next investigated whether TC input origin instructs distinct differentiation programs in VB-receiving (S1L4) and Po-receiving (S2L4)

Figure 2 | Po TC axons are rewired onto S1L4 neurons in vb^- mice.
a, **b**, Anterograde labelling in control mice showing arborization of VB axons within S1L4 (**a**, left) and exclusion of Po axons from this layer (**b**, left). In S2, VB axons are absent (**a**, right) whereas Po axons arborize in S2L4 (**b**, right). **c**, Po_{vb-} axons arborize within S1L4. **d**, Quantification. $**P < 0.005$, Student's t-test, $n = 3$ control (Ctrl) and $n = 4$ vb^- injections. **e**, Ultrasound (US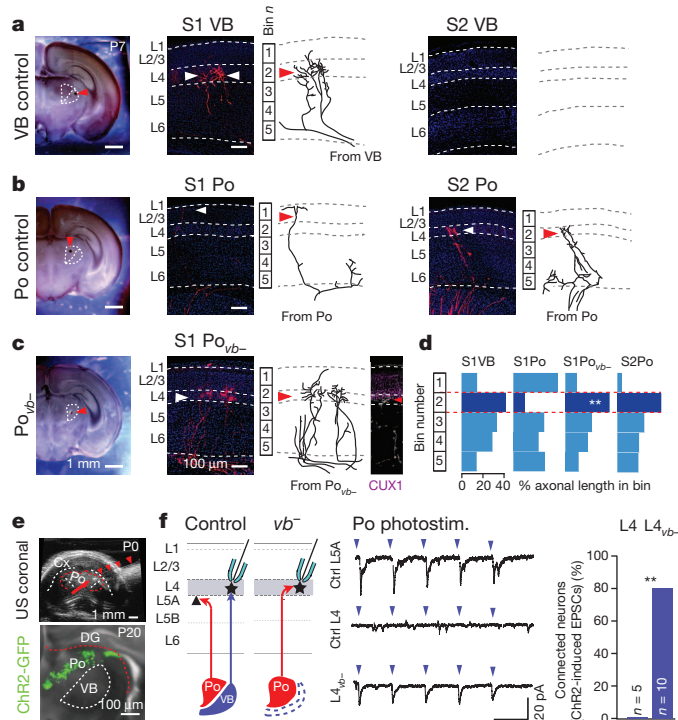)-guided microinjection of a AAV-ChR2-GFP virus into Po. Red arrowheads indicate location of the micropipette shaft. **f**, Optogenetic stimulation of Po axons elicits excitatory postsynaptic currents (EPSCs) in S1L5A neurons but not S1L4 neurons, whereas EPSCs are evoked in $S1L4_{vb-}$ neurons. $**P < 0.005$, Student's t-test.
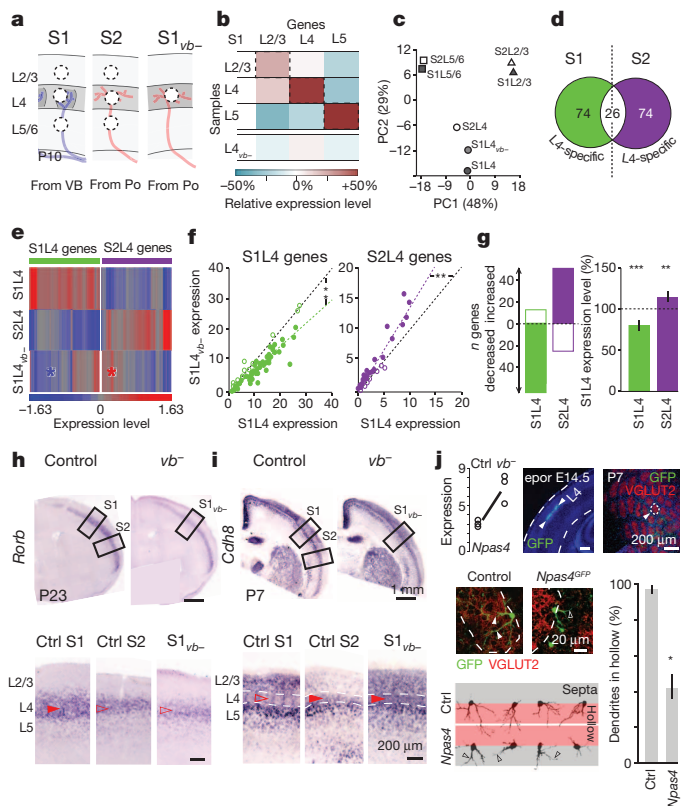
Figure 3 | VB→Po switch in input congruently respecifies $S1L4_{vb-}$ neurons. **a**, Schematic experimental design. **b**, Validation of sample collection with known marker genes for L2/3, L4 and L5/6 neurons (see ref. 22 and Methods). **c**, Principal component (PC) analysis plot of sample-specific gene expression showing clustering of L2/3 (triangles), L4 (circles) and L5/6 (squares) transcriptional programs. **d**, Delineation of S1L4-type and S2L4-type transcripts. **e**, Heatmap representation showing repression of a subset of S1L4-type genes (blue asterisk) and induction of S2L4-type genes (red asterisk) in $L4_{vb-}$ neurons. **f**, Levels of expression of S1L4-type genes (green: full, decreased; open, increased) and S2L4-type genes (purple: full, increased; open, decreased) in $L4_{vb-}$ neurons. $**P < 10^{-4}$, Student's t-test. **g**, 62 out of 74 S1L4-type genes are downregulated in $S1L4_{vb-}$ neurons; 50 out of 74 S2L4-type genes are upregulated in $S1L4_{vb-}$ neurons. $***P < 10^{-5}$, $**P < 0.005$, paired Student's t-test; $n = 74$ S1L4-type and 74 S2L4-type genes. **h**, **i**, In situ hybridization illustrating cell-type-specific downregulation of Rorb (**h**), and upregulation of Cdh8 (**i**). **j**, Overexpression of an $Npas4^{GFP}$-expressing construct into S1L4 neurons disrupts the polarization of dendrites towards barrel hollows (that is, VB axon-containing barrel centres) at P7. $*P < 0.05$, $n = 19$ Ctrl and $n = 10$ $Npas4^+$ neurons, ANOVA. Error bars denote mean ± s.e.m.

neurons. To examine whether the VB→Po switch in TC input repressed VB-target and induced Po-target differentiation programs, we first defined S1L4- and S2L4-type transcripts by identifying wild-type L4-specific genes whose expression was mutually exclusive between S1 and S2 (Fig. 3d). Analysis of the expression of these two gene sets by $S1L4_{vb-}$ neurons revealed a strong repression of S1L4-type genes and a corresponding induction of S2L4-type genes (Fig. 3e–g; $P < 0.0001$ for S1L4 ($n = 74$ genes) and < 0.005 for S2L4 ($n = 74$ genes) percentage expression level, paired Student's t-test). This respecification of molecular identity was further examined using in situ hybridization for select gene candidates including the S1L4-type transcript Rorb, which was strongly decreased in $S1L4_{vb-}$ neurons, and the S2L4-type transcript Cdh8, which was upregulated (Fig. 3h, i and Extended Data Fig. 4c–e). VB input-dependent changes in S1L4 gene expression coincided with the postnatal maturation of VB-L4 synapses, and failed to occur in $S1L4_{vb-}$ neurons, resulting in an S2L4-type developmental gene expression program (Extended Data Figs 1a–i, 5 and Supplementary Note 2).

We next investigated the respective contribution of VB input loss and Po input acquisition to these genetic changes. Comparison of vb^- mice

with *Dlx5/6::Cre/Celsr3*[flox] mice, which entirely lack TC input[23], revealed that each input type instructs the differentiation program of S1L4[vb−] neurons, as some genes were strictly VB-input-dependent, whereas others were Po-input-dependent (Extended Data Fig. 6 and Supplementary Note 3). Together, these data indicate that VB and Po axons exert input-type-specific controls over S1L4 neuron genetic identity.

We next examined the functional role of the top input-dependent S1L4 transcripts, which together formed a core set of genes involved in neurite differentiation and polarization (Extended Data Table 1). The transcription factors *Npas4* (ref. 24) and *Zbtb20* (ref. 25) and the extracellular protein *Sema3a* (ref. 26) were functionally characterized using *in vitro* and *in vivo* morphological analysis of embryonic day (E)14.5-electroporated L4 neurons. Each of these transcripts controlled neurite polarization *in vitro* and postnatal orientation of S1L4 neuron primary dendrites towards VB axons *in vivo*, a key requisite for columnar assembly of cortical circuits (Fig. 3j, Extended Data Fig. 7 and Supplementary Note 4). Thus, VB input exerts a genetic control over the development of critical morphological circuit-related properties of S1L4 neurons.

To further investigate the role of VB input in intracortical circuit assembly, we determined excitatory/inhibitory (E/I) feed-forward responses of L4 neurons to optogenetic stimulation of VB (for S1L4 neurons) and Po axons (for S2L4 and S1L4[vb−], Extended Data Fig. 8a). Supporting a TC-type-specific control over intracortical circuit assembly, S1L4[vb−] E/I responses were increased to S2L4 levels (Extended Data Fig. 8b–d; $P < 0.05$ for S1L4 versus S2L4 and S1L4 versus S1L4[vb−]; $P = 0.4$ for S2L4 versus S1L4[vb−], analysis of variance (ANOVA), $n = 5$ Ctrl, 7 S2L4 and 6 S1L4[vb−] neurons). VB→Po switch in input thus instructs corresponding downstream changes in intracortical circuit properties.

As VB neurons normally transmit tactile information[9,10] and Po neurons nociceptive information[12–14], we examined whether the VB→Po switch in input led to a congruent shift in the functional identity of S1L4 neurons (Fig. 4a). The trigeminal principalis nucleus (PrV), which provides input to VB, was notably atrophied in *vb−* mice, whereas Po afferent pathways remained unchanged, suggesting that the locus of the plasticity in *vb−* mice is essentially thalamocortical (Extended Data Fig. 9 and Supplementary Note 5). In control mice, single-whisker environmental exploration led to activation of S1L4 neurons in the corresponding

barrel, as reported by expression of the immediate-early gene *c-fos* (also known as *Fos*)[14,27] (Fig. 4b). S1L4[vb−] neurons were also activated by this task, although more diffusely, consistent with the broader distribution of presynaptic Po[vb−] terminals and S2-like increases in S1L4[vb−] E/I responses (Supplementary Note 6). Back-and-forth deflections of single whiskers in head-fixed mice evoked focal intrinsic optical signal responses in S1 (Fig. 4c)[28], which were weaker and more variable than in control mice, suggesting asynchronous cortical activation. Accordingly, *in vivo* intracellular recordings revealed a ~3-fold increase in the jitter of L2/3-evoked responses to whisker deflections and prolonged onset latencies in *vb−* mice (Fig. 4d, $P < 0.05$ for jitter and latency, Mann–Whitney rank-sum test, $n = 7$ Ctrl and 6 *vb−* cells). Consistent with this degradation in the coding of input signals, *vb−* mice were distinctly impaired at tasks requiring haptic skills and fine sensorimotor coordination (Fig. 4e and Supplementary Note 7). Together, these findings reflect the low fidelity of spatial and temporal encoding in Po ascending pathways[29] and indicate that Po-input-receiving S1L4[vb−] neurons still respond, albeit less reliably, to tactile stimuli *in vivo*.

To investigate whether respecified S1L4[vb−] neurons acquired the ability to respond to noxious stimuli, we performed a microinjection of capsaicin, an algogenic chemical, into the whisker pad of *vb−* mice[14]. In control mice, S1L5A neurons and S2L4 neurons were strongly activated, as revealed by c-Fos expression, while S1L4 neurons remained inactive, reflecting the normal target specificity of Po projections[14]. By contrast, S1L4[vb−] neurons were activated by noxious stimulation, in a pattern similar to that of S2L4 of control mice (Fig. 4f; $P < 0.001$ for L4 versus L4[vb−], Student's t-test; $n = 3$ Ctrl and $n = 3$ *vb−* mice). Consistent with an expanded L4 population response to Po input, *vb−* mice displayed increased behavioural signs of pain following capsaicin injection and shortened flick latencies to focal thermic tail stimulation (Supplementary Videos 1 and 2, Fig. 4g and Supplementary Note 7). Taken together, these data indicate that rewiring of Po input onto VB targets results in an abnormal convergence of normally segregated tactile and nociceptive sensory modalities onto S1L4 neurons.

Our findings reveal that distinct TC inputs exert modality-specific controls over the molecular identity and function of postsynaptic L4 neurons. VB and Po inputs only affected a specific subset of S1L4- and
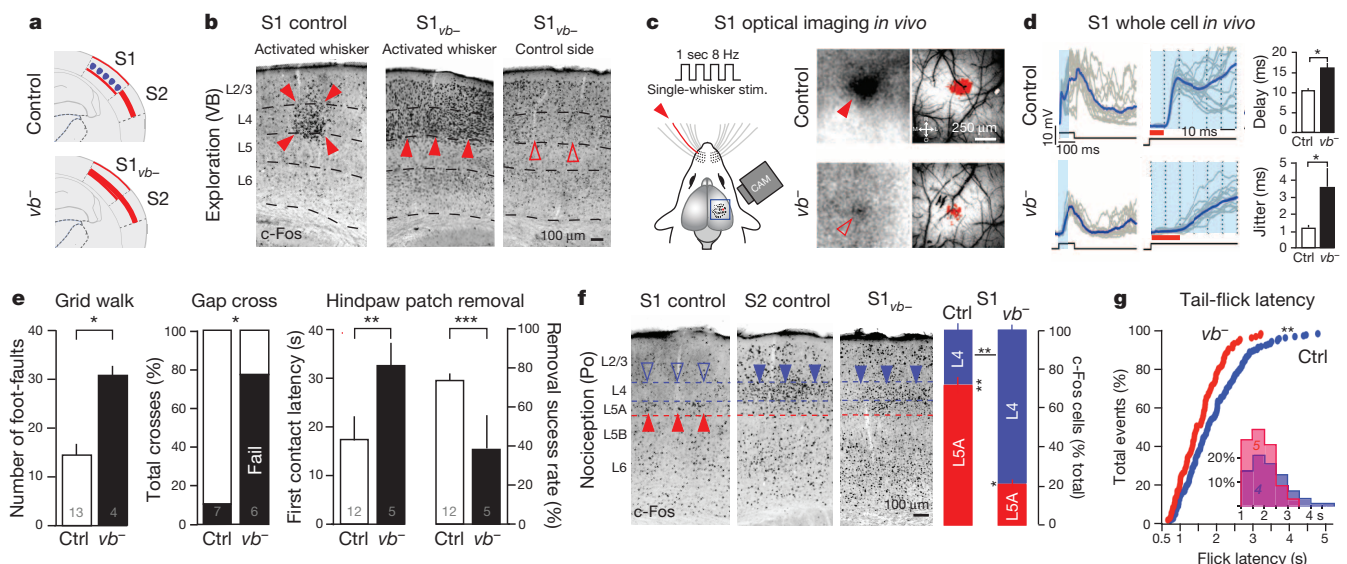
Figure 4 | Functional convergence of normally segregated sensory modalities in *vb−* mice. a, S1L4[vb−] neurons acquire S2L4-type functional properties. **b**, S1L4[vb−] neurons are activated during environmental exploration. **c**, Single-whisker stimulation elicits weak focal S1 responses in *vb−* mice. **d**, EPSCs from sample L2/3 neurons. Delay and increased jitter of responses to whisker deflections in *vb−* mice (*$P < 0.05$, Mann–Whitney rank-sum test, $n = 7$ Ctrl and 6 *vb−* neurons). **e**, Haptic skills and fine sensorimotor coordination are impaired in *vb−* mice (see Supplementary Note 7). *$P < 0.05$,

$P = 0.008$, ***$P < 0.0005$, Student's t-test. **f, Noxious stimulation of the whisker-pad activates S1L4[vb−] neurons but not S1L4 neurons ($n = 3$ Ctrl and $n = 3$ *vb−* mice; *$P < 0.05$, ** $P < 0.005$, Student's t-test) (L4 versus L5A in *vb−* mice, $P < 0.01$; L4 versus L5A in control mice, $P < 0.005$; L4 versus L4[vb−], $P < 0.005$). **g**, *vb−* mice have shorter flick latencies upon focal thermic stimulation of the tail. ** $P < 0.005$, Kolmogorov–Smirnov test. Error bars denote mean ± s.e.m.

S2L4-type genes, such that the identity of S1L4$_{vb-}$ neurons was intermediate between these two neuronal types. This intermediate genetic identity is reminiscent of the cytoarchitectonically 'hybrid' visual cortex found following *in utero* bilateral enucleation in monkeys[30], suggesting that input-type-dependent specification may have contributed to cortical neuron diversity during evolution.

Po-dependent 'non-lemniscal' somatosensory pathways are thought to be evolutionary older than VB-dependent 'lemniscal' pathways (see ref. 14 and references therein). Emergence of a specialized VB-dependent pathway during evolution suggests that VB axons were at competitive advantage over pre-existing Po axons. Supporting this possibility, our findings reveal that Po axons invade VB targets in vb^- cortex, and VB and Po axons co-mingle in the early postnatal cortex before segregating into their anatomically complementary patterns[16]. Therefore, evolutionary and developmental segregation of tactile and nociceptive pathways may have occurred via hierarchical interactions between VB and Po inputs.

Finally, the circuit-specific transcriptional controls described here provide a powerful mechanism matching terminal neuronal differentiation to specific functional constraints. Whereas cell-intrinsic differentiation programs initially define neuronal permissiveness to distinct inputs[15], reciprocally, these inputs differentially instruct gene expression programs in their targets. This crosstalk between gene expression and circuit connectivity may therefore act to orchestrate the assembly of cognate neurons into functionally specialized pathways during development.

METHODS SUMMARY

Animals. Compound mice were generated as described in Methods.

Histology. Brains were fixed and stained using standard methods[14,15].

Anterograde and retrograde labelling. Retrograde cortical labelling was performed using stereotaxic FluoroGold microinjections. Anterograde cortical labelling was performed on fixed brains using NeuroVue Dye Filter Red (MTTI, FS-1002) inserted into the VB or the Po.

Microdissection and microarray analysis. L2/3, L4 and L5/6 in S1 and S2 of P10 controls and vb^- mice were microdissected. After amplification and labelling of extracted RNA, hybridization onto Affymetrix Mouse Gene 1.0 ST arrays was performed.

ChR2 injections and electrophysiology. A ChR2-Venus-expressing adenovirus (AAV5-hSyn-hChR2 H134R) was used. Electrophysiology recordings were performed at 3 weeks of age[15].

In utero electroporations. These were performed as described in ref. 15.

Tactile and noxious stimulations. These were performed as detailed in ref. 14.

In vivo electrophysiology. Intrinsic imaging and whole-cell recordings were performed as detailed in ref. 28.

Behavioural tests. These were performed as described in the Methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Sur, M., Garraghty, P. E. & Roe, A. W. Experimentally induced visual projections into auditory thalamus and cortex. *Science* **242,** 1437–1441 (1988).
2. Dehay, C., Horsburgh, G., Berland, M., Killackey, H. & Kennedy, H. Maturation and connectivity of the visual cortex in monkey is altered by prenatal removal of retinal input. *Nature* **337,** 265–267 (1989).
3. Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274,** 1133–1138 (1996).
4. Miyashita-Lin, E. M., Hevner, R. F., Wassarman, K. M., Martinez, S. & Rubenstein, J. L. R. Early neocortical regionalization in the absence of thalamic innervation. *Science* **285,** 906–909 (1999).
5. Chou, S.-J. *et al.* Geniculocortical input drives genetic distinctions between primary and higher-order visual areas. *Science* **340,** 1239–1242 (2013).
6. Li, H. *et al.* Laminar and columnar development of barrel cortex relies on thalamocortical neurotransmission. *Neuron* **79,** 970–986 (2013).
7. Wimmer, V. C., Bruno, R. M., de Kock, C. P. J., Kuner, T. & Sakmann, B. Dimensions of a projection column and architecture of VPM and POm axons in rat vibrissal cortex. *Cereb. Cortex* **20,** 2265–2276 (2010).
8. Ohno, S. *et al.* A morphological analysis of thalamocortical axon fibers of rat posterior thalamic nuclei: a single neuron tracing study with viral vectors. *Cereb. Cortex* **22,** 2840–2857 (2012).
9. Nicolelis, M. A. L. Computing with thalamocortical ensembles during different behavioural states. *J. Physiol. (Lond.)* **566,** 37–47 (2005).
10. Yu, C., Derdikman, D., Haidarliu, S. & Ahissar, E. Parallel thalamic pathways for whisking and touch signals in the rat. *PLoS Biol.* **4,** e124 (2006).
11. Ahissar, E., Sosnik, R. & Haidarliu, S. Transformation from temporal to rate coding in a somatosensory thalamocortical pathway. *Nature* **406,** 302–306 (2000).
12. Gauriau, C. & Bernard, J.-F. Posterior triangular thalamic neurons convey nociceptive messages to the secondary somatosensory and insular cortices in the rat. *J. Neurosci.* **24,** 752–761 (2004).
13. Masri, R. *et al.* Zona incerta: a role in central pain. *J. Neurophysiol.* **102,** 181–191 (2009).
14. Frangeul, L. *et al.* Specific activation of the paralemniscal pathway during nociception. *Eur. J. Neurosci.* http://dx.doi.org/10.1111/ejn.12524 (2 March 2014).
15. De la Rossa, A. *et al. In vivo* reprogramming of circuit connectivity in postmitotic neocortical neurons. *Nature Neurosci.* **16,** 193–200 (2013).
16. Kichula, E. A. & Huntley, G. W. Developmental and comparative aspects of posterior medial thalamocortical innervation of the barrel cortex in mice and rats. *J. Comp. Neurol.* **509,** 239–258 (2008).
17. Viaene, A. N., Petrof, I. & Sherman, S. M. Properties of the thalamic projection from the posterior medial nucleus to primary and secondary somatosensory cortices in the mouse. *Proc. Natl Acad. Sci. USA* **108,** 18156–18161 (2011).
18. Jabaudon, D., Shnider, S. J., Tischfield, D. J., Galazo, M. J. & Macklis, J. D. *ROR*β induces barrel-like neuronal clusters in the developing neocortex. *Cereb. Cortex* **22,** 996–1006 (2012).
19. Narboux-Nême, N., Pavone, L. M., Avallone, L., Zhuang, X. & Gaspar, P. Serotonin transporter transgenic (SERTcre) mouse line reveals developmental targets of serotonin specific reuptake inhibitors (SSRIs). *Neuropharmacology* **55,** 994–1005 (2008).
20. Brockschnieder, D. *et al.* Cell depletion due to diphtheria toxin fragment A after Cre-mediated recombination. *Mol. Cell. Biol.* **24,** 7636–7642 (2004).
21. Endo, S. G-substrate: the cerebellum and beyond. *Prog. Mol. Biol. Transl. Sci.* **106,** 381–416 (2012).
22. Belgard, T. G. *et al.* A transcriptomic atlas of mouse neocortical layers. *Neuron* **71,** 605–616 (2011).
23. Zhou, L. *et al.* Early forebrain wiring: genetic dissection using conditional Celsr3 mutant mice. *Science* **320,** 946–949 (2008).
24. Yun, J. *et al.* Neuronal Per Arnt Sim (PAS) domain protein 4 (NPAS4) regulates neurite outgrowth and phosphorylation of synapsin I. *J. Biol. Chem.* **288,** 2655–2664 (2013).
25. Nielsen, J. V., Blom, J. B., Noraberg, J. & Jensen, N. A. Zbtb20-induced CA1 pyramidal neuron development and area enlargement in the cerebral midline cortex of mice. *Cereb. Cortex* **20,** 1904–1914 (2010).
26. Shelly, M. *et al.* Semaphorin3A regulates neuronal polarization by suppressing axon formation and promoting dendrite growth. *Neuron* **71,** 433–446 (2011).
27. Wagener, R. J., Dávid, C., Zhao, S., Haas, C. A. & Staiger, J. F. The somatosensory cortex of reeler mutant mice shows absent layering but intact formation and behavioral activation of columnar somatotopic maps. *J. Neurosci.* **30,** 15700–15709 (2010).
28. Gambino, F. & Holtmaat, A. Spike-timing-dependent potentiation of sensory surround in the somatosensory cortex is facilitated by deprivation-mediated disinhibition. *Neuron* **75,** 490–502 (2012).
29. Sosnik, R., Haidarliu, S. & Ahissar, E. Temporal frequency of whisker movement. I. Representations in brain stem and thalamus. *J. Neurophysiol.* **86,** 339–353 (2001).
30. Rakic, P., Suñer, I. & Williams, R. W. A novel cytoarchitectonic area induced experimentally within the primate visual cortex. *Proc. Natl Acad. Sci. USA* **88,** 2083–2087 (1991).

Author Contributions D.J. and G.P. conceived the project, D.J. and C.B. designed the electrophysiological experiments, and G.P., C.B., I.V., L.T., F.G. and D.J. performed the experiments. D.J. and G.P. wrote the manuscript with help from C.L. and A.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.J. (denis.jabaudon@unige.ch).

METHODS

Mice. C57BL/6 and CD1 mice, *Slc6a4*[Cre] transgenic mice[31], Ai14 transgenic reporter mice (Jackson Laboratories, stock number 007914)[32] and Rosa26::stop[flox]*DTA*[20] males and females were used. Experiments were carried out in accordance with the Institutional Animal Care and Use Committee of the University of Geneva and with permission of the Geneva cantonal authorities.

Histology. Postnatal mice were perfused with 4% paraformaldehyde (PFA) and brains were fixed overnight in 4% PFA at 4 °C. 50-μm vibratome sections (Leica VT1000S) were used for all histological experiments. S1L4 and S2L4 neurons were visually identified on the basis of tangential location using topographical atlases (S1 versus S2) and bin location (L4 = bins 3–4 out of 10, see Extended Data Fig. 2f, g).

In situ hybridization on slides was performed according to methods described previously[33]. In brief, hybridization was carried out overnight at 60 °C with the digoxigenin (DIG)-labelled RNA probes. After hybridization, sections were washed and incubated with alkaline phosphatase-conjugated anti-DIG antibody (Roche 1:2,000) overnight at 4 °C. After incubation, sections were washed and the colour reaction was carried out overnight at 4 °C in a solution containing NBT (nitro-blue tetrazolium chloride) and BCIP (5-bromo-4-chloro-3′-indoly phosphate *p*-toluidine salt) (Roche). After colour revelation, sections were washed, post-fixed for 30 min in 4% PFA and mounted with Fluromount (Sigma). For antisense probe synthesis, total complementary DNA was amplified by PCR with primers designed for specific messenger RNA sequence of *Pcdh20*, *NeuroD6* and *Grm4*. T7 or Sp6 promoter sequence was added to the reverse primer sequence. DIG-labelled antisense RNA probes were obtained after *in vitro* transcription of the resulted PCR product (Roche kit) or of the plasmid template for *Rorb* and *Cdh8* (kind gift from M. Studer).

For immunohistochemistry, brain sections were incubated 1 h at room temperature in a blocking solution containing 3% BSA and 0.3% Triton X-100 in PBS and incubated overnight at 4 °C with primary antibodies: rabbit anti-5HTT (1/500, Calbiochem, PC177L), rabbit anti-GSBS (1/2,000, gift from S. Endo)[21], rabbit anti-CALB2 (1/500, Swant, 7699/3H), guinea pig anti-VGLUT2 (1/2,000, Millipore, AB2251), mouse anti-GFAP (1/500, Sigma, G3893), rabbit anti-c-Fos (1/5,000, Santa Cruz, sc-52), rabbit anti-CUX1 (1:200, Santa Cruz, sc-13024). Sections were rinsed three times in PBS and incubated for 60–90 min at room temperature with the Alexa Fluor 488- or 546-conjugated secondary antibodies (1/500, Invitrogen). For c-Fos immunostaining, biotinylated goat secondary anti-rabbit antibody (1/200, Invitrogen) were used, followed by an amplification step with Vectastain ABC kit (Vector Laboratories) and revealed in a 0.05% DAB, 0.02% NiSO$_4$, 0.025% CoCl$_2$ and 0.01% H$_2$O$_2$ solution. DAPI (0.2 μg ml^{-1}) was used for fluorescent nuclear counterstaining.

For Nissl staining, brain sections were mounted, stained with 0.5% cresyl violet and dehydrated with graded alcohols. For cytochrome oxidase staining: free-floating sections were placed in a solution of 0.5 mg ml^{-1} DAB, 0.5 mg ml^{-1} cytochrome C (Sigma), 40 mg ml^{-1} sucrose, 0.1 mM Tris, pH 7.6 at 37 °C until staining appears. Fluoro-Jade B staining was performed as previously described[34] with Fluoro-Jade B kit (Millipore).

Imaging and quantifications. All photomicrographs were taken with an Eclipse 90i fluorescence microscope (Nikon, Japan) or with a LSM700 confocal microscope (Zeiss). Where cells were counted, this was performed on a minimum of three biological replicates, within a 0.25 mm^2 area encompassing the whole radial extent of the cortex, and divided into 5 or 10 bins. Photomicrographs are representative examples taken from ≥2 replicates.

Anterograde/retrograde labelling. Focal retrograde labelling from the cortex: anaesthetized P7 mice were placed on a stereotaxic apparatus thalamic neurons were retrogradely labelled via 55-nl injections of FluoroGold (FG) 2% (Hydroxystilbamidine bis(methanesulphonate), Fluorochrome Inc.). Brains were collected 2 days after the injection. Focal retrograde labelling from the Po (Extended Data Fig. 9d): 92 nl of Alexa Fluor 555-conjugated cholera toxin subunit B (Invitrogen) was microinjected stereotactically at P10 and brains were collected at P23 and sectioned coronally. Only injections confined to the Po were analysed (*n* = 4 Ctrl and *n* = 3 *vb*$^-$ injections).

For anterograde labelling, the caudal part of fixed P7 brains was cut to reveal the caudal thalamus. Trypan blue staining was applied to allow distinction between thalamic nuclei and anterograde labelling was initiated by insertion of small NeuroVue Dye-coated Filter Red (MTTI, FS-1002) (50 × 50 μm^2) in VB or Po. Brains were incubated at 37 °C in 0.4% PFA for 8 weeks and cut into 100-μm sections on a vibratome before immediate imaging. TC layer-specific arborization was quantified by measuring axonal length within each of the 5 bins using ImageJ software. Analysis was performed on 3–5 100-μm-thick sections corresponding to bregma levels −0.82 to −1.58 on the Paxinos adult brain atlas; S1 analyses were performed in the cortex at the same dorsolateral level as the hippocampus, whereas S2 was identified as the region lateral to this, adjacent to the piriform cortex.

Tissue microdissection and microarray. One mouse of control and *vb*$^-$ littermates was used to collect each of three biological replicates of L2/3, L4, L5/6 samples at P10 for cortical microarrays, or of Po and VB for thalamic microarrays.

Fresh coronal brain sections (140 μm) were cut on a vibrating microtome and thalamic nuclei or cortical layers were visually identified and microdissected using a Leica Stereomicroscope (Leica M165FC) in ice-cold oxygenated artificial cerebrospinal fluid under RNase-free conditions. Samples were stored in RNAlater at −80 °C. RNA was extracted using an RNeasy kit (Qiagen) and one- or two-cycle amplification and labelling was performed (according to Affymetrix protocols) using Superscript cDNA synthesis kit (Invitrogen), MEGAscript T7 kit and MessageAmp IIaRNA amplification kit (Ambion). Labelled cRNA was fragmented and hybridized to Affymetrix Mouse Gene 1.0 ST arrays (for cortical samples) or mouse 430 2.0 Genome arrays (for thalamic samples). GeneChips were incubated at 45 °C for 16 h with biotin-labelled cRNA probes, and then washed and stained using a streptavidin-phycoerythrin conjugate with antibody amplification as described in Affymetrix protocol, using Affymetrix GeneChip Fluidics Station 450. GeneChips were scanned on a GCS3000 scanner (Affymetrix).

Microarray CEL files were normalized using Robust Multichip Analysis and analysed using Partek Genomics Suites software (http://www.partek.com). The accuracy of the microdissection approach was validated by measuring concordance of gene expression of the samples with genes confirmed to be strongly enriched in one cortical layer. Relative gene expression = (expression in the defined layer – mean expression in all layers)/(mean expression in all layers)[22]. We defined the laminar identity of L4, L2/3 and L5/6 neurons within S1 and S2 by using ANOVA (fold change >2; *P* < 0.05, adjusted for multiple comparisons within Partek Genomic Suites) to identify their top 100 most specifically expressed genes in each of the six collected control samples. The relatedness between these six samples and L4$_{vb-}$ was determined using unbiased cluster analysis of this reference set of genes (total 600 transcripts, of which 453 are unique).

In utero electroporation and tissue culture. E14.5 timed pregnant CD1 mice were anaesthetized with isoflurane/oxygen and a pCAG-*GFP* reporter plasmid or CMV-*SEMA3A-Tomato*[35] (kind gift from F. Ango), pCAG-*Npas4-GFP* (Thermo Fisher Scientific, MMM1013-202733015) or pCAG-miR*Zbtb20-GFP*[25] (kind gift from N. A. Jensen) construct was injected into the embryos' lateral ventricle using a 40-μm-tip glass micropipette mounted on a Nanoject II nanoinjector (Drummond Scientific). E14.5 was chosen as a gestational age to specifically target L4 neuron progenitors[15]. Voltage pulses (40 mV, 50 ms) were applied using external paddles in order to target S1, as described previously[15]. Females were allowed to give birth and P7 pups were fixed with intracardiac perfusion of 4% PFA. Tangential sections of flattened electroporated cortex were cut at 50 μm and immunostaining for VGLUT2 was performed. Images were taken at LSM700 Zeiss confocal. The delineation between barrel septae and hollows was performed on serially reconstructed image stacks using the ImageJ software. The total number of primary dendrites and distribution of dendrites within hollows or septae were determined and statistically analysed using an ANOVA test across conditions.

For *in vitro* culture, electroporated mouse embryos were dissected at E16.5 in ice-cold HBSS. The S1-positive electroporated site was microdissected under fluorescent microscopy and dissociated cells cultured during 3 days on 12-mm, coated coverslips (poly-L-lysine, 0.1 mg ml^{-1}) in 500 μl Neurobasal supplemented with GlutaMAX, B27, sodium pyruvate and a mixture of antibiotic penicillin-streptomycin all from Gibco. The cultures were then fixed with 4% PFA, mounted with DAPI and analysed on a LSM 700 Zeiss confocal and ImageJ software.

In vitro electrophysiology. AAV-mediated expression of ChR2 (AAV5-hsynhChR2 H134R) was injected at P0 in the Po, under ultrasound guidance (Vevo 660, VisualSonics). Injected mice were collected 3 weeks later and processed for electrophysiology. ~P23 mice were deeply anaesthetized with isoflurane and were then decapitated. Brains were removed and placed in cold (0–4 °C), oxygenated (95% O$_2$–5% CO$_2$) slicing solution containing: 119 mM NaCl, 2.5 mM KCl, 1.3 mM MgCl, 2.5 mM CaCl$_2$, 1.0 mM Na$_2$HPO$_4$, 26.2 mM NaHCO$_3$ and 11 mM glucose. Coronal slices (300 mM) were kept at room temperature and were allowed to recover for at least 1 h before recording. Under low magnification, the barrels in L4 could be readily identified, and high-power magnification was used to guide the recording electrode onto visually identified neurons. The radial extent of the cortex was virtually divided into 5 bins (1 = most pial) and patched neurons were always located in bin 2, which corresponds to L4, as validated in pilot experiments in which neurons were filled with biocytin (3 mg ml^{-1}) for 15 min and location was assessed using immunostaining of CUX1 after fixing fresh sections (*n* = 7 Ctrl and *n* = 6 *vb*$^-$) in both S1 and S2. The internal solution contained 140 mM potassium gluconate, 5 mM KCl, 10 mM HEPES, 0.2 mM EGTA, 2 mM MgCl$_2$, 4 mM Na$_2$ATP, 0.3 mM Na$_3$GTP and 10 mM sodium creatine-phosphate. Currents were amplified (Multiclamp 700B, Axon Instruments), filtered at 5 kHz and digitized at 20 kHz (National Instruments Board PCI-MIO-16E4, Igor, WaveMetrics). The liquid junction potential was +12 mV. Experiments were discarded if the access resistance varied by more than 20%. Synaptic currents were evoked by light flashes delivered by a fibre optic cable (Thorlabs) attached to a 473-mm solid-state laser (CrystaLaser). Cells were held at −60 mV and GABA currents were blocked by wash in picrotoxin

(Tocris, 100 μM). Light-evoked EPSCs were recorded at −46 mV and light-evoked IPSCs were recorded at 0 mV (ref. 36) (Extended Data Fig. 8).

In vivo electrophysiology

Intrinsic optical imaging. ~P23 mice were first anaesthetized with isoflurane (4% for induction with ~0.5 l min^{-1} O$_2$), and then with urethane (1.5 g kg^{-1}, i.p., prepared in lactated ringer solution containing in mM: 102 NaCl, 28 Na L-Lactate, 4 KCl, 1.5 CaCl$_2$). Eye ointment was applied to prevent dehydration. The scalp was locally anaesthetized with lidocaine (1%), the periosteum gently removed, and a custom-made plastic chamber was attached to the skull above barrel cortex (centred 1.5–2 mm posterior from bregma, 2–2.5 mm lateral) with dental acrylic and dental cement. The chamber was filled with sterile cortex buffer (containing in mM: 125 NaCl, 5 KCl, 10 glucose, 10 HEPES, 2 CaCl$_2$, and 2 MgSO$_4$, pH 7.4) and sealed with a glass coverslip. Intrinsic optical signals were imaged through the intact skull using an Imager 3001F[28].

In vivo whole-cell patch clamp. After imaging, a small, ~1 × 1 mm piece of bone was removed using a dental drill (centred above the C2 whisker maximum intrinsic optical signal response). Whole-cell 'blind' patch-clamp recordings were obtained as previously described[28]. High positive pressure (200–300 mbar) was applied to the pipette (7–9 MΩ) to prevent tip occlusion while penetrating the dura. After passing the dura the positive pressure was immediately reduced to prevent cortical damage. The pipette was then advanced in 2-μm steps, and pipette resistance was monitored in the conventional voltage-clamp configuration. When the pipette resistance suddenly increased, positive pressure was relieved to obtain a 3–5-GΩ seal. After break-in, the membrane potential (V_m) was measured, and dialysis was allowed to occur for at least 5 min before deflecting the whisker. Data were acquired using a Multiclamp700B amplifier (Molecular Devices), and digitized at 10 kHz (National Instruments), using MATLAB (Mathworks)-based Ephus software (http://research.janelia.org/labs/display/ephus; The Janelia Farm Research Center). Off-line analysis was performed using custom routines written in IgorPro (Wavemetrics). All neurons were located at ~200 μm below the pia.

Current-clamp recordings were made using a potassium-based internal solution (in mM: 135 potassium gluconate, 4 KCl, 10 HEPES, 10 Na$_2$-phosphocreatine, 4 Mg-ATP, 0.3 Na-GTP, pH adjusted to 7.25 with KOH, 285 mOsm). Series resistance (R_s) and input resistance (R_{in}, not including R_s) were monitored with a 100-ms long-lasting hyperpolarizing square pulse 400 ms before each whisker deflection, and extracted off-line by using a double exponential fit. Recordings were discarded if one of the following conditions occurred: (1) V_m and R_s exceeded −50 mV and 50 MΩ, respectively; (2) spontaneously occurring spikes were not overshooting; (3) R_s or R_{in} changed more than 30% over the duration of the experiment. The bridge was usually not balanced and liquid junction potential was not corrected.

Whisker-evoked post-synaptic potential (PSP) analysis. Whisker-evoked PSPs were evoked by forth and back deflection of the whisker (100 ms, 0.1 Hz) using piezoelectric ceramic elements attached to a glass pipette ~4 mm away from the skin. The voltage applied to the ceramic was set to evoke a whisker displacement of ~0.6 mm with a ramp of 7–8 ms. The C1 and C2 whiskers were independently deflected by different piezoelectric elements. PSP analyses were confined to down states. Peak amplitude and integral analysis was performed on each trace, and then presented as a mean of at least 30 whisker-evoked responses. The PSP onset latency was defined as the time point at which the amplitude exceeded 3 × s.d. of the baseline noise over 5 ms before stimulation. The jitter was defined as the standard deviation of at least 30 whisker-evoked responses (control mice, $n = 7$ cells per 3 mice; vb^- mice, $n = 6$ cells per 3 mice).

Environmental enrichment. All but one whisker (C1) on the right side of the snout were clipped in P23 vb^- or control mice ($n = 3$), after which the animals were placed in an large playground box containing plastic balls, maze-like pieces of thread, and various small objects, as previously described. Mice were kept for 1 h in enriched environment, perfused and brains stained for c-Fos expression[14,27,37].

Noxious stimulation. P23 vb^- or control mice were briefly anaesthetized with isoflurane, and 50 μl of a capsaicin solution (10 mM (Sigma), 100% ethanol and 7% Tween-80 in saline) was subcutaneously injected in the whisker pad as previously described[14,38]. Mice were allowed to wake up and were euthanized after 1 h and their brain stained for c-Fos expression. The number of c-Fos$^+$ neurons was quantified on five sections per animal ($n = 3$) in a ~600-μm width of S1 (~3 barrels). Statistical comparisons of layer-specific c-Fos activation between control and vb^- mice were done using Student's t-test.

Behavioural analysis. ~P30 vb^- mice and littermates were used for all behavioural experiments. Both males and females were used. To habituate animals to the testing environment, mice were transported within their home cage to the testing room for three consecutive days before testing. All behavioural tests took place during the light phase of the light–dark cycle, with the observer blinded to the mouse genotype.

For the open field test, spontaneous locomotor activity was examined in $n = 6$ vb^- mice and $n = 17$ control littermates. Mice were placed in the centre of a 45 × 45-cm arena and locomotion in the dark (distance travelled, percentage time spent in central, intermediate or lateral sectors) was recorded during 15 min using an infrared camera coupled to a tracking software (ANYmaze, Stoelting Co). For the grid-walk test, sensorimotor abilities (limb placement accuracy, coordination) of hindlimbs were examined in $n = 4$ vb^- mice and $n = 13$ control littermates by assessing the aptitude of the mice to navigate in dim lighting over a wire mesh grid (2.4 × 2.4-cm grid spaces, 45 × 45-cm total area) during 6 min[39,40]. Foot-faults were counted when a hindlimb paw protruded entirely through the grid. For the gap-crossing test, whisker-sensing abilities and sensorimotor coordination were examined in $n = 7$ vb^- mice and $n = 6$ control littermates by placing mice on a 12 × 12 cm elevated platform and measuring the number of crosses performed to a neighbouring platform across a 4-cm gap in obscurity. The gap was such that the mouse was required to extend its head and detect the opposite platform with its whiskers before crossing[41,42]. The number of crosses and failed attempts (that is, falls) were recorded by an observer via an infrared camera. A maximum of two falls per mouse was admitted after which the test was interrupted. Adhesive patch removal task: this test was originally developed to assess somatosensory asymmetry and sensory function after sensorimotor cortex lesions[43–45]. $n = 5$ vb^- mice and $n = 12$ control littermates were used. A 6-mm diameter circular adhesive patch was placed on the plantar surface of each hindpaw, after which mice were released in the testing arena and observed for 240 s. Latency to detect the first patch (snout contact with the patch) as well as the time taken to remove both patches was measured. Mice underwent three consecutive trials, with an inter-trial interval of 60 min, and values were averaged for each mouse. For the tail-flick test, sensitivity to noxious stimuli was determined using the tail-flick test[46] in $n = 5$ vb^- mice and $n = 4$ control littermates. In the apparatus (IITC Life Science), a pre-focused light beam supplied an area of 4 × 6-mm heat stimulation to the tail. The time taken for the mouse to flick its tail away from the stimulation area was recorded, providing a measure of pain sensitivity. Values were obtained from three sessions of ten trials (inter-session interval: 24 h).

Statistics. No statistics were used to determine group sample size; however, sample sizes were similar to those used in previous publications from our group and others. The person performing the test was blinded to the animal's genotype. All mice were used in the study; the tail-flick test was performed only if the animal collaborated to rest with its tail in the flick detection groove. If animals undertook more than one task, the order of the task was randomized. Two-tailed t-tests were used for all statistical analyses except for the tail-flick test, for which values had a non-Gaussian distribution, which was analysed using a Kolmogorov–Smirnov test. Values are shown as mean ± s.e.m. throughout the manuscript. n values refer to biological replicates throughout the manuscript.

31. Zhuang, X., Masson, J., Gingrich, J. A., Rayport, S. & Hen, R. Targeted gene expression in dopamine and serotonin neurons of the mouse brain. *J. Neurosci. Methods* **143,** 27–32 (2005).
32. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13,** 133–140 (2010).
33. Lai, T. *et al.* SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. *Neuron* **57,** 232–247 (2008).
34. Schmued, L. C. & Hopkins, K. J. Fluoro-Jade B: a high affinity fluorescent marker for the localization of neuronal degeneration. *Brain Res.* **874,** 123–130 (2000).
35. Cioni, J.-M. *et al.* SEMA3A signaling controls layer-specific interneuron branching in the cerebellum. *Curr. Biol.* **23,** 850–861 (2013).
36. Gabernet, L., Jadhav, S. P., Feldman, D. E., Carandini, M. & Scanziani, M. Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. *Neuron* **48,** 315–327 (2005).
37. Bisler, S. *et al.* Expression of c-Fos, ICER, Krox-24 and JunB in the whisker-to-barrel pathway of rats: time course of induction upon whisker stimulation by tactile exploration of an enriched environment. *J. Chem. Neuroanat.* **23,** 187–198 (2002).
38. Noma, N. *et al.* Organization of pERK-immunoreactive cells in trigeminal spinal nucleus caudalis and upper cervical cord following capsaicin injection into oral and craniofacial regions in rats. *J. Comp. Neurol.* **507,** 1428–1440 (2008).
39. Z'Graggen, W. J., Metz, G. A., Kartje, G. L., Thallmair, M. & Schwab, M. E. Functional recovery and enhanced corticofugal plasticity after unilateral pyramidal tract lesion and blockade of myelin-associated neurite growth inhibitors in adult rats. *J. Neurosci.* **18,** 4744–4757 (1998).
40. Leingärtner, A. *et al.* Cortical area size dictates performance at modality-specific behaviors. *Proc. Natl Acad. Sci. USA* **104,** 4153–4158 (2007).
41. Hutson, K. A. & Masterton, R. B. The sensory contribution of a single vibrissa's cortical barrel. *J. Neurophysiol.* **56,** 1196–1223 (1986).
42. Jenkinson, E. W. & Glickstein, M. Whiskers, barrels, and cortical efferent pathways in gap crossing by rats. *J. Neurophysiol.* **84,** 1781–1789 (2000).
43. Hernandez, T. D. & Schallert, T. Seizures and recovery from experimental brain damage. *Exp. Neurol.* **102,** 318–324 (1988).
44. Thallmair, M. *et al.* Neurite growth inhibitors restrict plasticity and functional recovery following corticospinal tract lesions. *Nature Neurosci.* **1,** 124–131 (1998).
45. Tomassy, G. S. *et al.* Area-specific temporal control of corticospinal motor neuron differentiation by COUP-TFI. *Proc. Natl Acad. Sci. USA* **107,** 3576–3581 (2010).
46. Le Bars, D., Gozariu, M. & Cadden, S. W. Animal models of nociception. *Pharmacol. Rev.* **53,** 597–652 (2001).
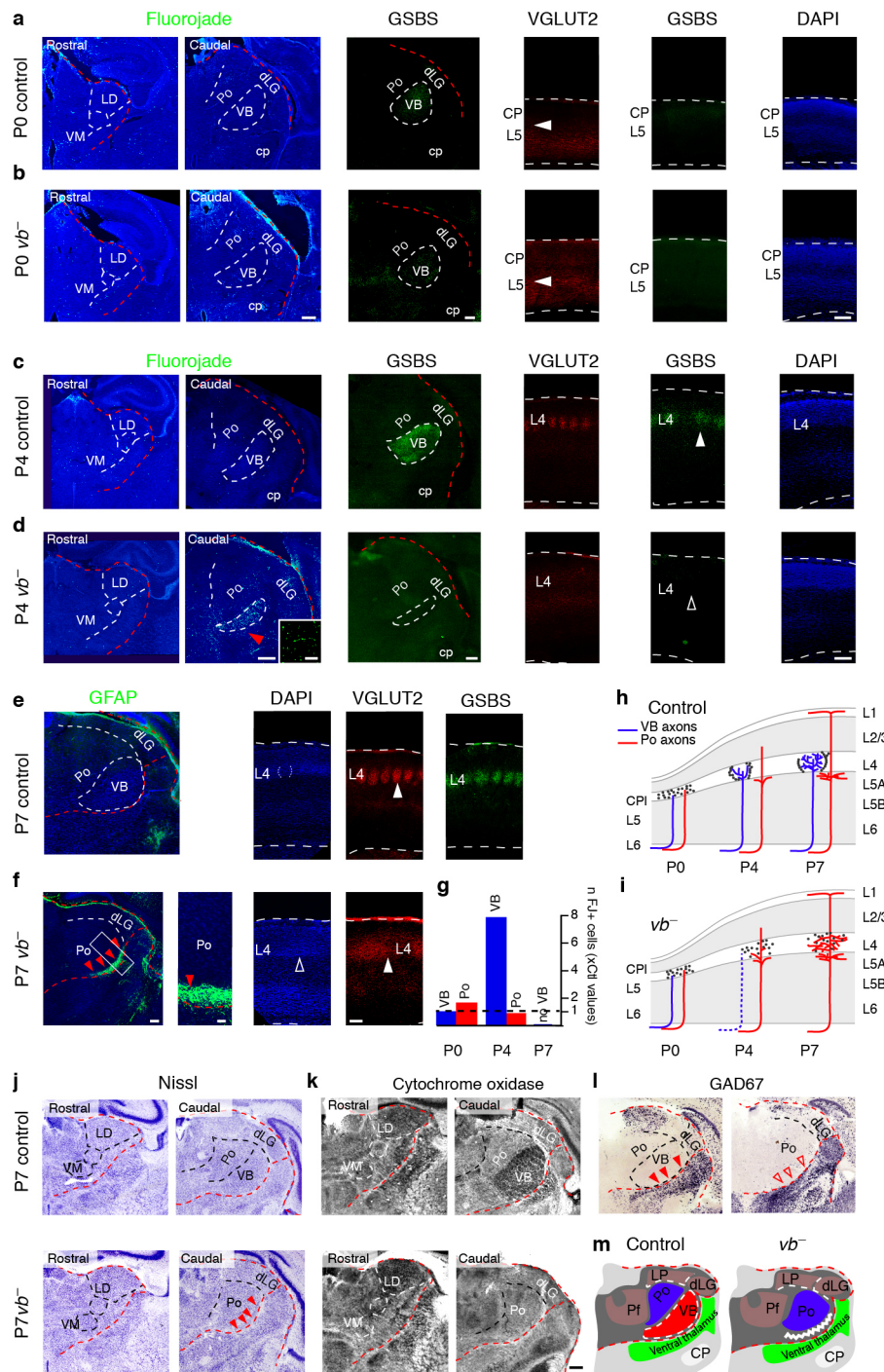
Extended Data Figure 1 | Early postnatal VB ablation in *vb⁻* mice.
a–f, Coronal sections of the thalamus and S1 cortex at P0 (**a**, **b**), P4 (**c**, **d**) and P7 (**e**, **f**) in control (**a**, **c**, **e**) and *vb⁻* (**b**, **d**, **f**) mice. Staining for Fluoro-Jade B (FJ) reveals a peak of VB-specific neuronal degeneration at P4 in *vb⁻* mice, a time at which the VB-specific marker GSBS[21] is already lacking. VGLUT2 immunostaining indicates that TC axons are present in deep cortical layers of *vb⁻* mice at P0 (arrowhead) and reach L4 at P4, a time at which VB degeneration is at its peak. These axons are not clustered into whisker-related barrels and do not express the VB-specific marker GSBS, which is normally expressed at P4. DAPI (4′,6-diamidino-2-phenylindole) staining indicates that postsynaptic L4 neurons fail to assemble into barrels in *vb⁻* cortex. At P7,

immunostaining for the glial marker GFAP shows a secondary glial 'scar' (red arrowheads) at the ventral thalamic border in *vb⁻* mice (**e**, **f**). **g**, Quantification of FJ⁺ neurons. **h**, **i**, Schematic representation of VB (blue) and Po (red) axonal development in control (**h**) and *vb⁻* (**i**) S1 cortex. **j**, **k**, Nissl (**j**) and cytochrome oxidase (**k**) stainings reveal specific ablation of the VB in *vb⁻* mice. **l**, *In situ* hybridization for the interneuron marker GAD67 shows preserved delineation of the dorsal and the ventral thalamus (red arrowheads). **m**, Schematic representation of *vb⁻* thalamus. Scale bars, 200 μm (insets, 50 μm). CP, cerebral peduncle; CP, cortical plate; dLG, dorsolateral geniculate nucleus; LD, laterodorsal nucleus; LP, lateroposterior nucleus; PF, parafascicular nucleus; VM, ventromedial nucleus.
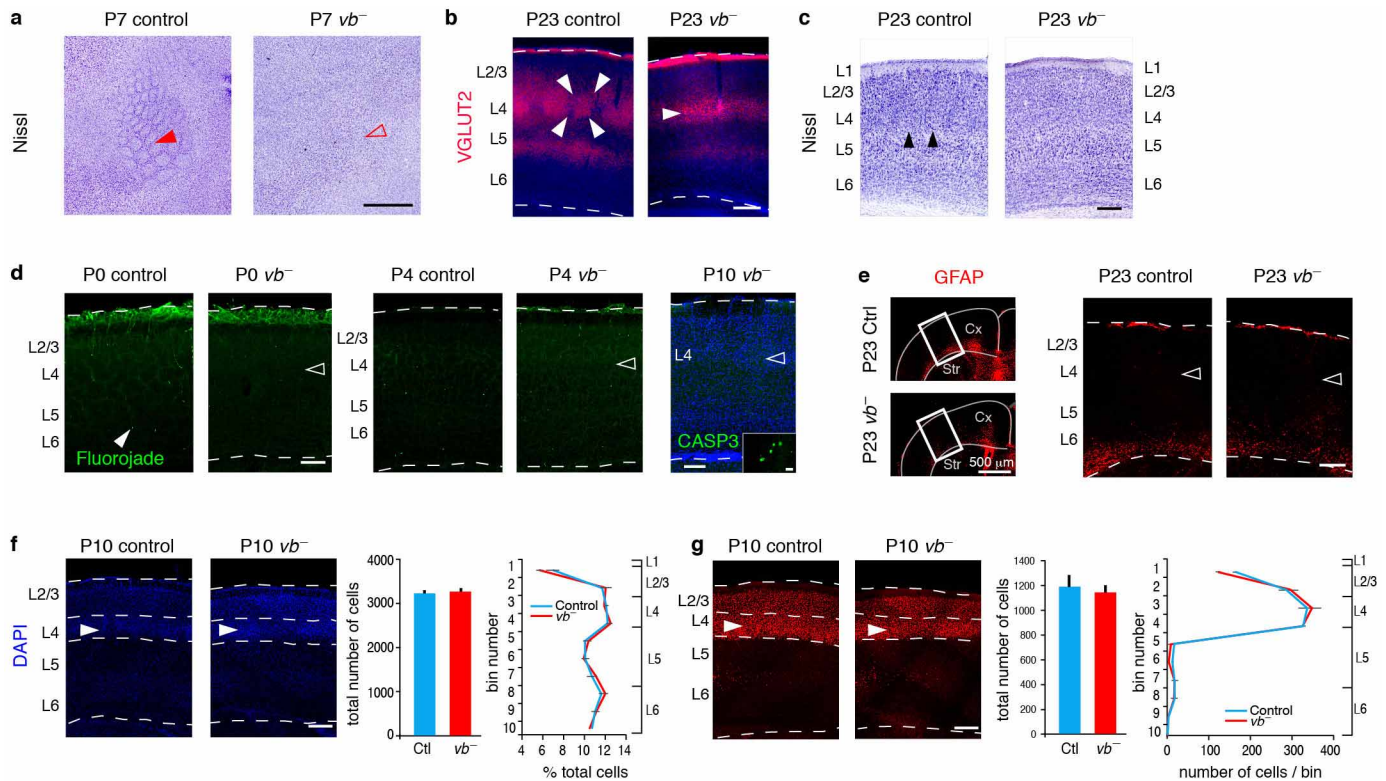
Extended Data Figure 2 | Rewiring of input to S1L4 occurs in the absence of cortical neuron death. a, Nissl-stained flattened preparation of the somatosensory cortex showing lack of barrel-like clusters in postsynaptic L4 cortical neurons in vb^- mice (empty red arrowhead). **b, c**, Presynaptic TC terminals are still present in L4 of vb^- cortex at P23 using the pan-TC presynaptic marker VGLUT2 (white arrowheads) (**b**) whereas cortical barrels are lacking (**c**). **d, e**, Fluoro-Jade B, cleaved caspase 3 (CASP3) (**d**) and GFAP (**e**) expression do not show evidence of cortical neuron degeneration in vb^- mice (inset shows non-cortical CASP3$^+$ neurons in a non-cortical region of the same section). **f, g**, Staining and quantification using DAPI (**f**) and the L2/3-L4 marker CUX1 (**g**) show lack of barrels but preservation of S1L4 cell numbers in vb^- mice. Total quantification surface: 0.25 mm². Scale bars, 400 μm (**a**), 200 μm (**b–d, f, g**), inset 10 μm (**d**).
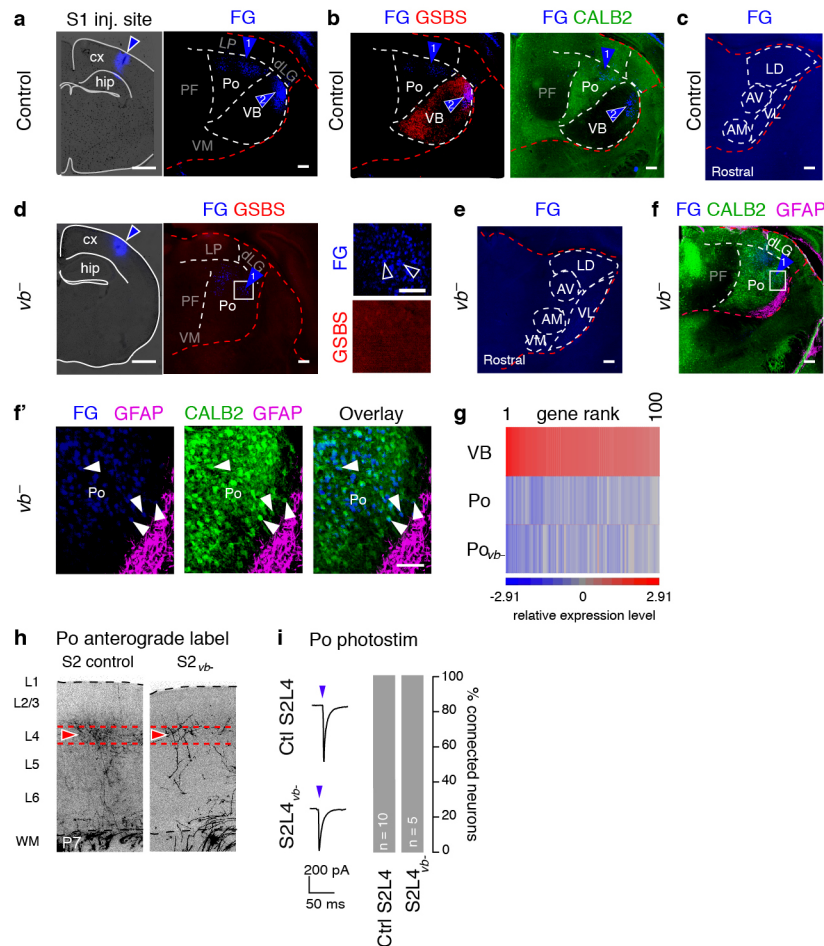
Extended Data Figure 3 | Ectopic TC projections to S1 in *vb*⁻ mice originate from the Po. **a–c**, In control mice, retrograde labelling from S1 using FluoroGold (FG) labels both GSBS⁺ VB neurons and CALB2⁺ Po neurons (**b**, numbered blue arrowheads), whereas other thalamic nuclei are not detectably labelled (**b**, **c**). **d–f′**, In *vb*⁻ mice, retrograde labelling from S1 exclusively labels GSBS⁻ Po neurons (**d**, blue arrowhead). No additional labelling is found in other thalamic nuclei (**d**, **e**). FG⁺ Po neurons are CALB2⁺ and located outside of the glial scar (**f**, **f′**: high magnification from inset in **f**). **g**, Po_{vb−} neurons were undistinguishable from control Po neurons by molecular and microarray comparative gene expression analysis between Po_{vb−}, Po and VB neurons demonstrating that they are bona fide Po neurons. Heatmap representation of the expression intensity of the 100 most VB-specific genes in VB, Po and Po_{vb−} neurons. None of these genes are statistically significantly upregulated in Po_{vb−} neurons compared to Po. **h**, **i**, Po-S2L4 connectivity is normal in *vb*⁻ as assessed by anterograde labelling of Po projections in S2L4_{vb−} cortex (**h**) and S2L4 neuron responses to optogenetic stimulation of Po axons (**i**). Scale bars, 1 mm (low-magnification images) (**a–f**) and 100 μm elsewhere. AM, anteromedial; AV, anteroventral; Cx, cortex; dLG, dorsolateral geniculate; Hip, hippocampus; LD, laterodorsal; LP, lateroposterior; PF, parafascicular; VL, ventrolateral; VM, ventromedial.
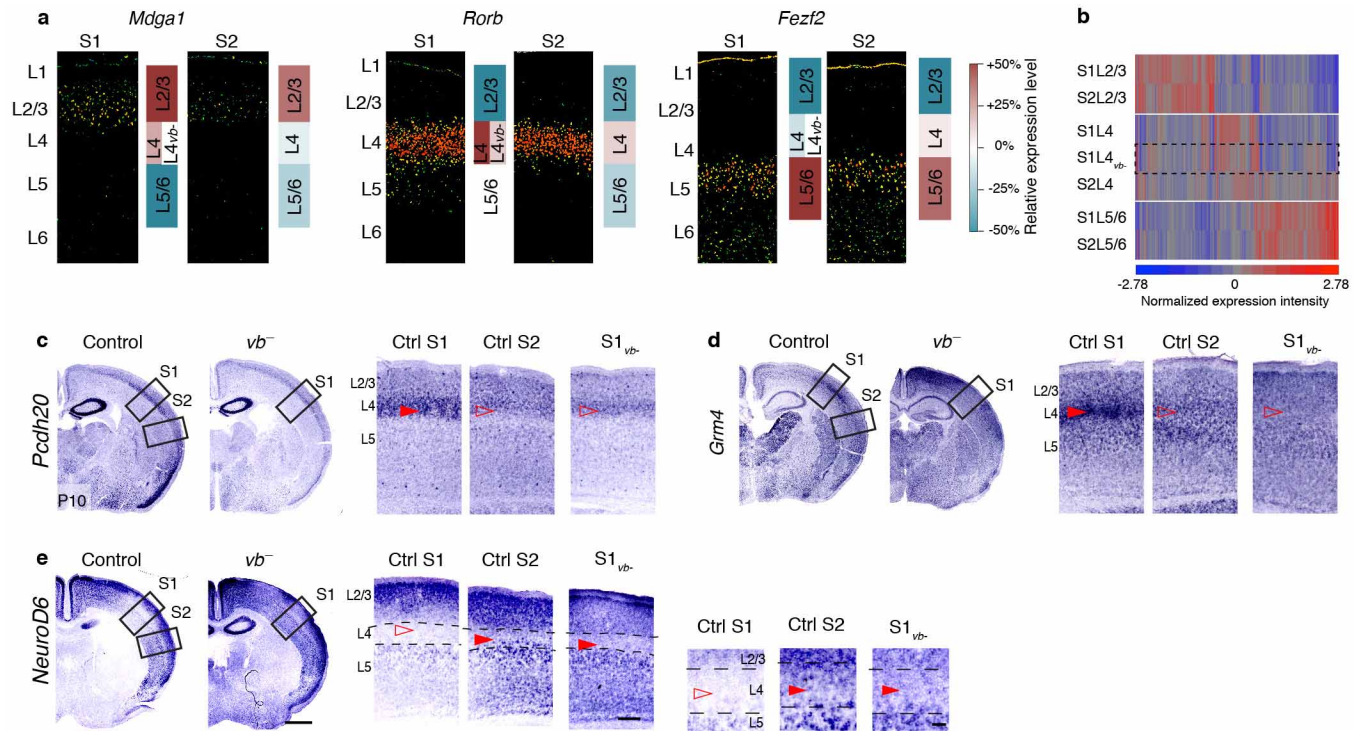
Extended Data Figure 4 | VB→Po switch in input leads to downregulation of S1L4 transcripts and upregulation of S2L4 transcripts in S1L4$_{vb-}$ neurons. a, *In situ* hybridizations (expression density map) from the Allen Brain Atlas (ABA) database showing expression of three sample genes specifically expressed in L2/3, L4 and L5 (*Mdga1*, *Rorb*, *Fezf2*, respectively). The heatmap on the right represents relative expression intensity in the microarray samples, which is concordant with the ABA data. Relative expression = (expression in the defined region – mean expression in all layers)/

(mean expression in all layers) (see ref. 22). **b**, Heatmap representing sample-specific gene expression for the union of the top 100 most specifically expressed genes of L4$_{vb-}$, L2/3, L4 and L5/6 samples in S1 and S2. Note that S1L4$_{vb-}$ gene expression is intermediate between S1L4 and S2L4 neurons. **c–e**, *In situ* hybridization showing downregulation of the S1L4-enriched transcripts *Pcdh20* (**c**) and *Grm4* (**d**), and upregulation of S2L4-enriched transcripts *NeuroD6* (**e**) in S1L4$_{vb-}$ neurons. Scale bars, 1 mm (low magnifications), 100 μm (high magnifications), 30 μm (inset).
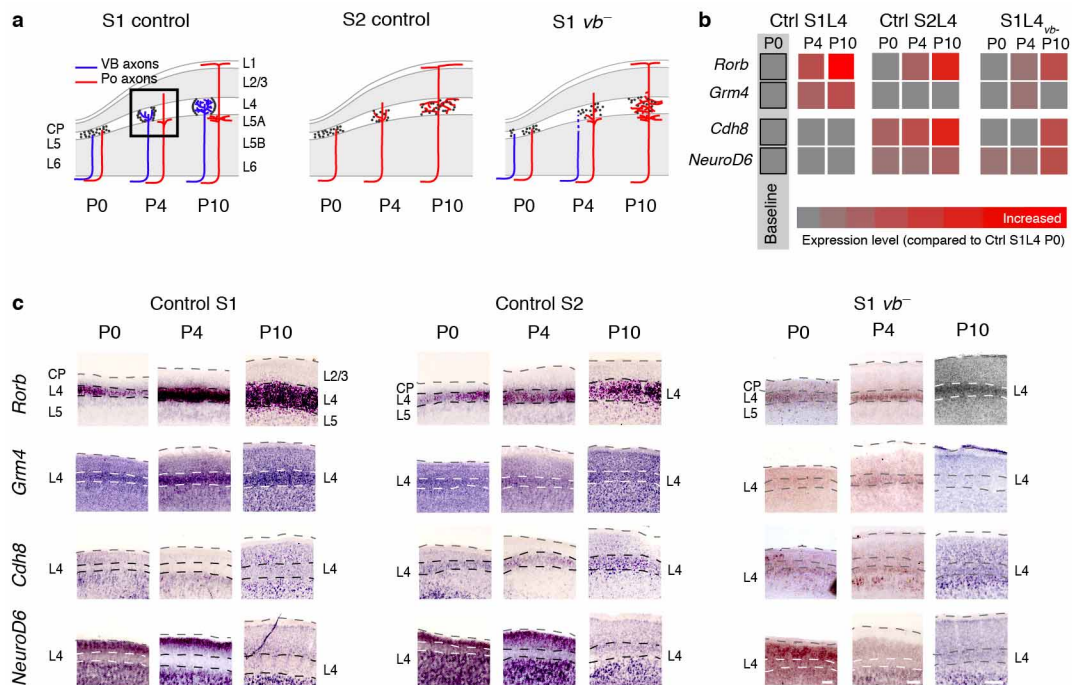
Extended Data Figure 5 | S1L4$_{vb-}$ developmental gene expression is S2L4-like. a, Schematic representation of VB (blue) and Po (red) axonal development in control S1 (left), control S2 (middle) and S1$_{vb-}$ (right) cortex. Boxed area indicates region shown in **c**. **b**, Summary of the findings: the time course of expression of L4 gene expression in S1L4$_{vb-}$ neurons is similar to that of S2L4 control cortex. Values are colour-coded using S1L4 P0 control values as baseline. **c**, *In situ* hybridizations for *Rorb* and *Grm4* (S1L4-type) and *Cdh8* and *NeuroD6* (S2-L4-type) transcripts at P0, P4 and P10 indicate that S1L4$_{vb-}$ developmental gene expression is S2L4-like. Scale bars, 100 μm. CP, cortical plate. See also Supplementary Note 2.
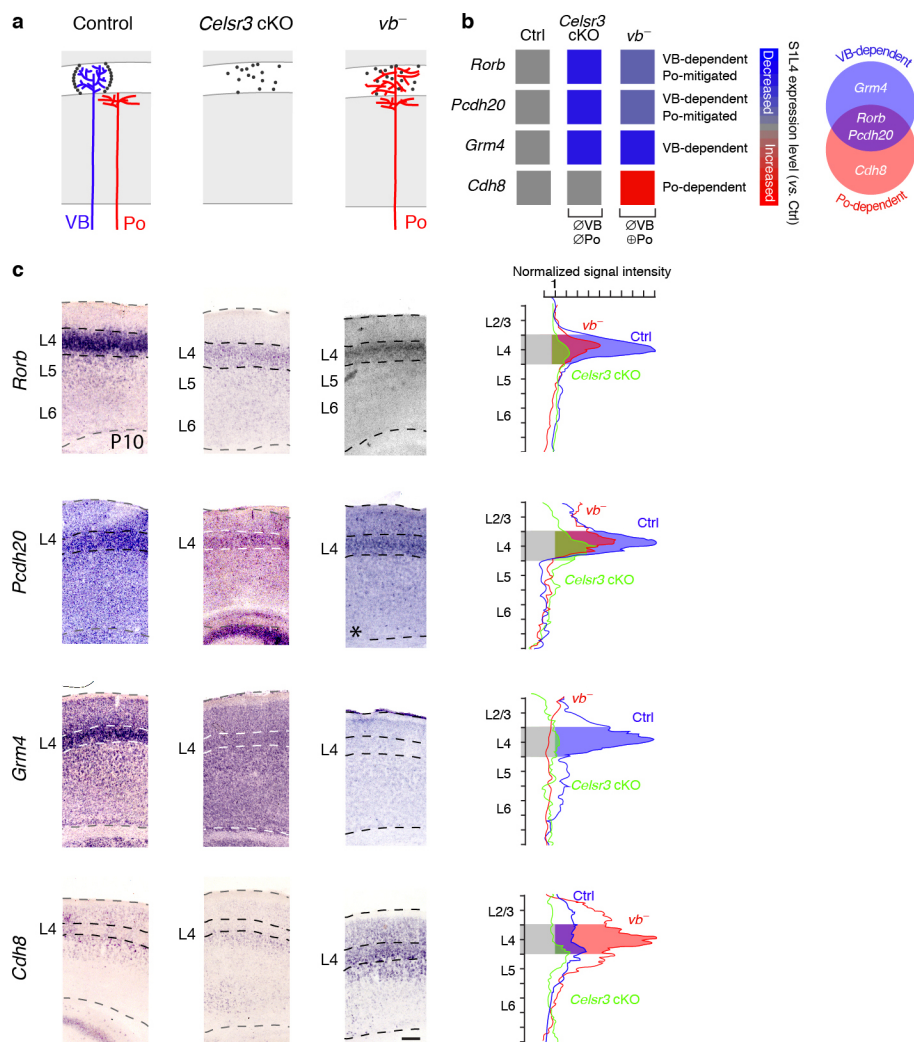
Extended Data Figure 6 | Loss of VB input and acquisition of Po input each define genetic changes in S1L4$_{vb-}$ neurons. a, Schematic representation of the phenotypes examined. *Celsr3* conditional knockouts (cKO) S1L4 neurons lack both VB and Po inputs (see Supplementary Note 3) whereas S1L4$_{vb-}$ receive Po but not VB input. **b**, Summary of the findings: expression of *Rorb* and *Pcdh20* expression is decreased both in *Celsr3* cKO and *vb*$^-$ L4 neurons, but this decrease is mitigated by Po input in *vb*$^-$ cortex. By contrast, *Grm4*

expression is not rescued by Po input and is thus VB-dependent. *Cdh8* upregulation in *vb*$^-$ L4 neurons depends on Po input as it does not occur in the absence of TC input (*Celsr3* cKO). **c**, *In situ* hybridizations showing expression of the S1L4-type genes *Rorb*, *Pcdh20* and *Grm4*, and the S2L4-type gene *Cdh8*. The two photomicrographs with an asterisk are also presented in Extended Data Fig. 4. Normalized intensity values were obtained by radial scanning of intensity using the gel tool of ImageJ software. Scale bars, 200 µm.
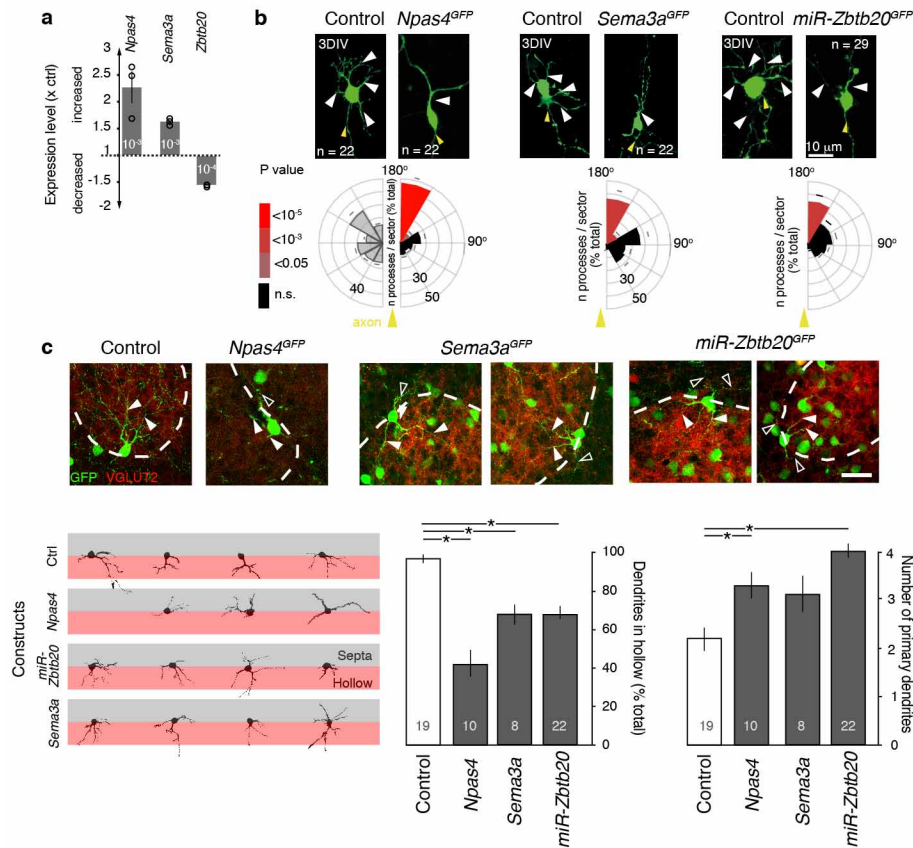
Extended Data Figure 7 | VB input regulates expression of genes controlling neurite differentiation and polarity in S1L4 neurons. **a**, *Npas4* and *Sema3a* expression is increased in *vb*⁻ mice, whereas *Zbtb20* is decreased. Open circles indicate values for individual replicates, value within bars indicate *P* value obtained with the microarray analysis. **b**, Overexpression of *Npas4* and *Sema3a* or downregulation of *Zbtb20* using a miR construct in S1L4 neurons collected 2 days after *in utero* electroporation at E14.5 led to changes in cell polarity. Dendrites are preferentially oriented away from the axon (yellow arrowhead, 0° in bulls eye plot), whereas they are evenly distributed when a control^GFP plasmid is used. *n* and *P* values are indicated in the figure, ANOVA. **c**, *In vivo* overexpression of *Npas4*, *Sema3a* or *miR-Zbtb20* by *in utero* electroporation at E14.5, the time of birth of L4 neurons, impairs dendritic orientation towards VB axons located in barrel hollows at P7 (VGLUT2⁺ region within barrels, pink area in single-cell displays) and increases in the number of primary dendrites for *Npas4* and *miR-Zbtb20* (*P < 0.05, ANOVA, *n* values indicated within bars). Septae correspond to spaces between barrels (grey area in single cell displays). Scale bars, 10 μm (**b**), 20 μm (**c**). See also Supplementary Note 4. Values: mean ± s.e.m.
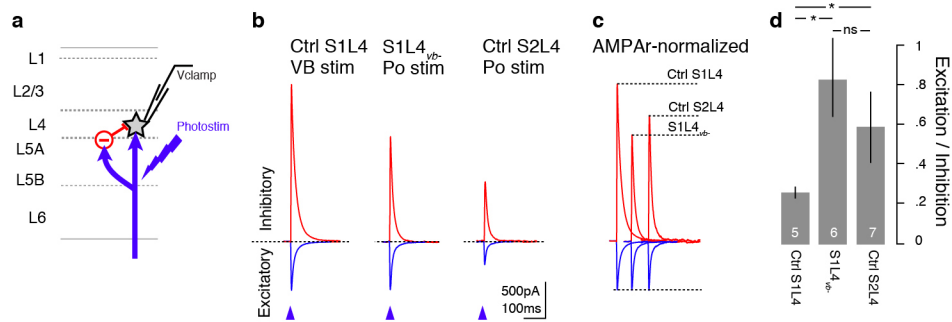
Extended Data Figure 8 | S1L4$_{vb-}$ E/I input balance is S2L4-like.
a, Schematic summary of the experiment: S1L4, S2L4 and S1L4$_{vb-}$ neurons were recorded while optogenetically stimulating the VB (for S1L4 neurons) or the Po (for S2L4 and S1L4$_{vb-}$ neurons). Feed-forward inhibitory input onto L4 neurons was determined by changing the holding potential as detailed in Methods. **b–d**, Sample traces (**b**) normalized to the excitatory input amplitude (**c**) and average values (**d**) showing that E/I ratios are increased to S2L4 levels in S1L4$_{vb-}$ neurons. Values: mean ± s.e.m. *$P < 0.05$, ANOVA; ns, not significant ($P = 0.4$).
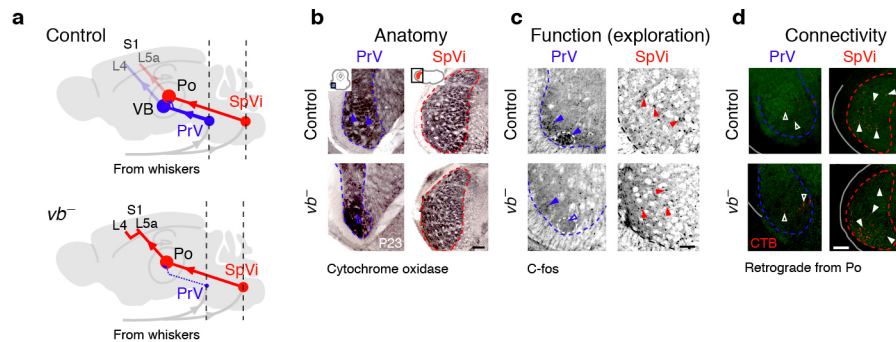
Extended Data Figure 9 | Pre-thalamic trigeminal input pathways are not rewired in *vb⁻* mice. a, Schematic representation of trigeminothalamic pathways and summary of the findings. Input to the VB, which conveys information on whisker contacts, originates from the PrV nucleus of the trigeminal complex, forming the lemniscal pathway, with a small contingent of fibres reaching the VB–Po border, see Supplementary Note 6 for details). Input to the Po, which forms the paralemniscal pathway, originates from the SpVi nucleus (interpolaris part of the spinal nucleus) of the trigeminal complex. In *vb⁻* mice, the PrV nucleus is markedly atrophied, presumably owing to loss of VB targets, and only a few cells subside. The paralemniscal pathway is unaffected. Dashed lines indicate location of the sections shown in **b–d**. **b**, Cytochrome oxidase staining showing markedly atrophied PrV in *vb⁻* mice. **c**, The PrV is not detectably activated by whisker contacts during environmental exploration in *vb⁻* mice, whereas the SpVi is unaffected. **d**, Retrograde labelling from the Po shows numerous labelled neurons in the SpVi and sparse labelled neurons in both control and *vb⁻* mice (*n* = 5 Ctrl and *n* = 3 *vb⁻* injections). Scale bars, 100 µm.

Extended Data Table 1 | VB input regulates a core set of genes involved in neurite differentiation and polarization

Gene symbol	Refseq ID	Fold-change (*vb vs* WT)	Function	Reference (Pubmed ID)
Nrn1	**NM_153529**	**2.93**	**extracell prot., dendritic growth**	**23115177**
NeuroD6	NM_009717	2.55	TF, axonal elongation	23303943
Nurr77	NM_010444	2.45	nuclear receptor, neurite extension	20375114
Vgf	**NM_001039385**	**2.39**	**extracell prot., dendritic growth**	**23115177**
<u>Npas4</u>	NM_153553	2.23	TF, neurite outgrowth	23172225
Frzb	NM_011356	2.21	extracell prot., axonal growth	16172602
Camk1g	NM_144817	2.19	prot. kinase, axonogenesis	19657032
Rorb	**NM_001043354**	**-2.17**	**nuclear receptor, adhesion**	**21799210**
N-Shc	NM_009167	2.15	TrkA receptor BP, neurite extension	12446789
Pcdh20	NM_178685	-2.02	adhesion molecule, dendritic self-avoidance	22842903
Unc5d	AK035883	-1.95	receptor, multipolar->radial polarity	22726835
Vmat2	**NM_172523**	**-1.91**	**vesicular monoamine transporter**	**12435414**
Trim67	NM_198632	-1.88	ubiquitin ligase, neuritogenesis	22337885
Rgs4	NM_009062	1.84	G-protein-binding, axonogenesis	23052218
EphA8	NM_007939	-1.80	Receptor, neurite outgrowth	15782114
Lmo4	**NM_010723**	**1.93**	**TF, Ca^{2+}-dependent**	**16899735**
Spatial	NM_001017433	-1.80	kinesin-binding prot., neurite outgrowth	24361585
Cdh18	NM_001081299	1.76	cadherin family member	17133224
Ocam	NM_001113208	1.75	membrane prot., neurite polarity	16531066
Robo3	NM_001164767	1.73	receptor, axonal guidance	18466743
ROLP	NM_183428	-1.67	adhesion molecule	19564939
<u>Sema3a</u>	NM_009152	1.63	extracell prot., neuronal polarization	21835341
PlxnD1	NM_026376	-1.62	Sema receptor, synaptic specificity	23395374
Cntn4	NM_001109749	1.62	membrane prot, axonal guidance	18367085
Fat3	NM_001080814	-1.59	membrane receptor, neurite pruning	21903076
Elmo1	NM_080288	1.58	spine morphogenesis	21900250
Prg5	NM_029425	1.56	membrane protein, neurite outgrowth	20032306
Cdh20	NM_011800	1.53	cadherin family member	17133224
Btbd3	**NM_145534**	**-1.52**	**TF, dendritic polarization**	**24179155**
Bhlhb5	**NM_021560**	**-1.52**	**TF, postsynaptic barrel assembly**	**18957218**
Odz3	NM_011857	-1.52	membrane prot, neuritogenesis	22367537
<u>Zbtb20</u>	NM_019778	-1.51	TF, cortical arealization, dendritic tufting	23283686, 19955470
Flrt3	NM_001172160	1.51	Unc5 binding prot., axonal guidance	21673655, 24560577

Genes with a demonstrated role in S1 barrel cortex formation, S1L4 neuron polarization, or TC circuit assembly appear in bold. The three genes functionally investigated in the current study are underlined. Negative values in fold-change indicate that expression is decreased in S1L4$_{vb-}$ neurons compared to control.

LETTER

Receptor binding by H10 influenza viruses

Sebastien G. Vachieri[1]*, Xiaoli Xiong[1]*, Patrick J. Collins[1], Philip A. Walker[1], Stephen R. Martin[1], Lesley F. Haire[1], Ying Zhang[1], John W. McCauley[1], Steven J. Gamblin[1] & John J. Skehel[1]

H10N8 follows H7N9 and H5N1 as the latest in a line of avian influenza viruses that cause serious disease in humans and have become a threat to public health[1]. Since December 2013, three human cases of H10N8 infection have been reported, two of whom are known to have died. To gather evidence relating to the epidemic potential of H10 we have determined the structure of the haemagglutinin of a previously isolated avian H10 virus and we present here results relating especially to its receptor-binding properties, as these are likely to be major determinants of virus transmissibility. Our results show, first, that the H10 virus possesses high avidity for human receptors and second, from the crystal structure of the complex formed by avian H10 haemagglutinin with human receptor, it is clear that the conformation of the bound receptor has characteristics of both the 1918 H1N1 pandemic virus[2] and the human H7 viruses isolated from patients in 2013 (ref. 3). We conclude that avian H10N8 virus has sufficient avidity for human receptors to account for its infection of humans but that its preference for avian receptors should make avian-receptor-rich human airway mucins[4] an effective block to widespread infection. In terms of surveillance, particular attention will be paid to the detection of mutations in the receptor-binding site of the H10 haemagglutinin that decrease its avidity for avian receptor, and could enable it to be more readily transmitted between humans.

Sequence analysis of the haemagglutinin (HA) of a recent human isolate of H10N8 virus[1] indicates its close similarity to an avian H10 HA that we have selected for detailed analysis of receptor-binding avidity and specificity (Extended Data Fig. 1). We use our data on the avian H10 virus to deduce the properties of the recent H10N8 virus isolated from infected humans. Similar approaches have provided mechanistic comparisons of the HAs of H1 (refs 5, 6), H2 (refs 5, 7) and H3 (ref. 8) avian viruses that are potential precursors of the 1918 H1, 1957 H2 and 1968 H3 pandemic viruses, and, most recently, of the HAs of H5 (refs 9, 10) and H7 (refs 11, 12) avian viruses related to the H5 and H7 viruses isolated from infected humans. We find that avian H10, like H5 and H7 viruses, has a preference for avian as opposed to human receptor and that, of the receptor analogues tested, it binds sulphated avian receptors most tightly.

We used biolayer interferometry to measure the binding of the avian H10 virus to human and avian receptor analogues (Fig. 1a). For comparison, we include in this figure data for human and avian H7 viruses binding to the same receptors[3]. The data show, first, that the avian H10 virus binds to human receptors substantially tighter than avian H7 virus and somewhat tighter even than human H7 virus. Second, avian H10 binds avian receptor significantly tighter than either human or avian H7 viruses. This is also the case for sulphated receptor analogues, a property that has been suggested to correlate with an ability to infect humans[13]. We find that H10 binds sulphated 3′-sialyl-N-acetyllactosamine (3′-SLN) most tightly of the receptor analogues examined, 3–5 times tighter than 3′-SLN (Extended Data Fig. 2). A comparison of receptor binding by avian H10 and two pandemic viruses—H1 (1918 Spanish influenza) and H3 (1968 Hong Kong influenza)—shows that human receptor binding by avian H10 virus is almost as strong as that of either pandemic virus (Fig. 1b). However, the preference of the pandemic viruses for human

versus avian receptors (10- to 100-fold) is not shared by the avian H10 virus which, in contrast, binds avian receptor about 150-fold tighter than human receptor. Preference of avian H10 virus for avian versus human receptor is consistent with its prevalence in birds and suggests that efficient infection of humans would require mutations that cause a
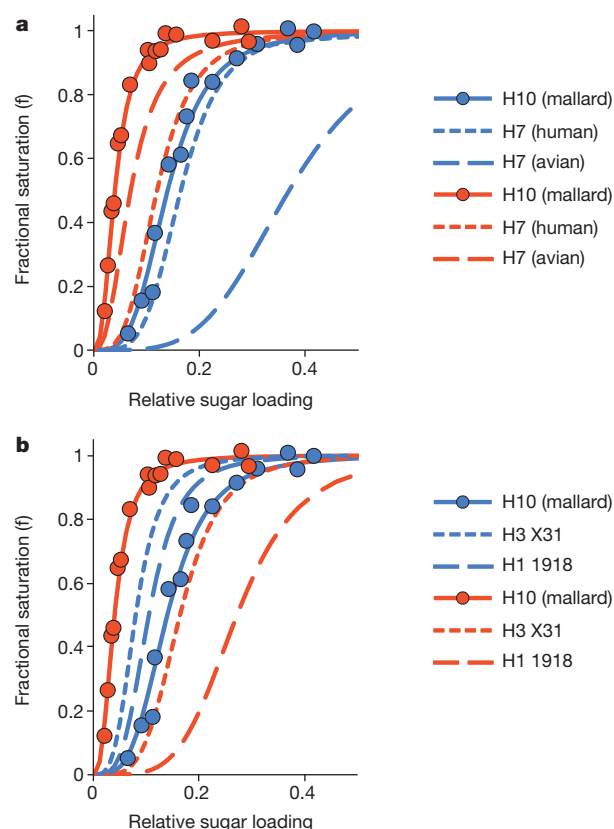
Figure 1 | Receptor binding to H10. The binding of influenza virus to avian (3′-SLN, red symbols and lines) and human (6′-SLN, blue symbols and lines) receptor analogues monitored by biolayer interferometry[26]. The graphs show fractional saturation of the sensor surface as a function of relative sugar loading for a fixed virus concentration of 100 pM. **a**, Data for the binding of avian H10N2 virus are shown in comparison to binding of human and avian H7 viruses. The data show that the avian H10 virus binds to human receptors substantially more tightly than avian H7 virus (∼ 100-fold) and somewhat more tightly even than human H7 virus (∼twofold). Also, avian H10 virus binds avian receptor significantly more tightly than either human (∼150-fold) or avian H7 (∼tenfold) viruses. **b**, H10N2 virus binding relative to two pandemic viruses H1N1 (1918) and H3N2 (1968). Human receptor binding to avian H10 virus is almost as strong (5–20-fold weaker) as that of both of the pandemic, and thus human-adapted, viruses. However, the preference of the pandemic viruses for human versus avian receptors (approximately 10–20-fold for H3 and ∼100-fold for H1) is reversed in the avian H10 virus which binds avian receptor about 150-fold more tightly than human.

[1]MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.
*These authors contributed equally to this work.

shift towards preference for human receptor. This is because avian-receptor-rich mucins in the human respiratory tract competitively inhibit infection by viruses that prefer to bind avian receptors[4]. Furthermore, viruses that prefer avian receptors replicate less efficiently in human respiratory epithelial cells[14]. Pandemic viruses have a marked preference for human over avian receptor binding and the probable mutations that reduce avian receptor binding, and result in this reversal of preference, have been characterized. Thus, for the H1 pandemic virus, evolution from its potential avian precursor primarily involved a mutation in the 190-helix of the receptor-binding site, Glu190Asp[2] and for the H2 and H3 pandemic viruses, the potential avian precursors required a Gln226Leu substitution in the 220-loop[15,16].

H10 HAs are members of the same genetic clade as H7 HAs and a comparison, by X-ray crystallography, of the structure of the avian H10 HA with the HA from an H7 virus isolated from humans[3] shows that they share a number of structural features (Extended Data Fig. 1, crystallographic statistics in Extended Data Table 1). Of particular note are the clade-specific sharp turn at the amino terminus of α-helix B and the elongated 150-loop (Extended Data Fig. 3) that protrudes into the receptor-binding site[3]. The locations of the differences in amino acid sequence between the avian H10 (ref. 17) and the recently isolated human H10 (ref. 1) are also shown in Extended Data Fig. 1.

The amino acid sequences of the receptor-binding sites of the avian H10 HA and the HA of the recent H10 isolate from humans are very similar. Residues that are known to influence receptor binding in other HAs, such as residues 186 and 190 at the N terminus of the 190-helix, and residues 223, 225 and 226 of the 220-loop, are conserved between them. Only residue 137 at the carboxy terminus of the 130-loop differs by the conservative substitution Lys137Arg. There is, therefore, a strong likelihood that our observations on the receptor-binding properties of avian H10 predict the properties of the H10 virus isolated from humans in China. However, although Lys 137 in avian HA makes no contact with receptors, Arg 137 in the human HA has the potential to make a number of interactions, as we describe below.

To investigate the structural basis of receptor-binding specificity and affinity we determined the structure of avian H10 HA in complex with both human and avian receptor analogues (Fig. 2) and human H10 in complex with human receptor (Extended Data Fig. 1b), and compared them with equivalent complexes formed by HAs from pandemic[5,8] and H7 (ref. 3) viruses. Electron density for human receptor bound to avian H10 HA is strong for the sialic acid (Sia-1) and galactose-2 (Gal-2) components of the receptor, and defined, but somewhat weaker, for N-acetyl glucosamine-3 (NAG-3). The receptor adopts a *cis* conformation about the Sia-1–Gal-2 glycosidic bond, as observed in human receptor complexes
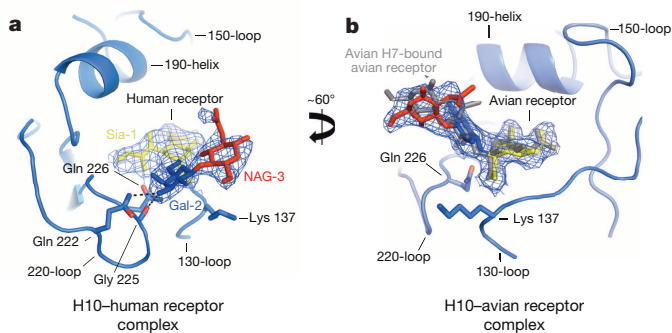
formed with pandemic viruses[5,8]. However, there is a notable difference in the angle of rotation around the Gal-2 C6–C5 bond of the receptor between the H10 complex and the complexes formed by HAs of pandemic viruses (Fig. 3). In the case of H10 the sialosaccharide exits the binding pocket, over the 130-loop in a trajectory less vertical than that seen in H1–receptor and H3–receptor complexes but very similar to that taken by the human receptor in the H7–receptor complex (Fig. 3).

Of the HAs of pandemic viruses, only H1 retains the characteristic avian glutamine residue at position 226 (ref. 18). The avian H10 HA also contains a glutamine residue at this position. Pandemic H1 HA binds human receptor with Gln 226 positioned lower in the base of the receptor pocket compared with its position when avian receptor is bound. As a result, Gal-2 of the human receptor is also lower in the binding site, closer to the 130-loop, than in the human receptor complexes formed by the HAs of H2 and H3 pandemic viruses, both of which contain a hydrophobic leucine residue at this position. Gln 226 in H10 also sits lower in the receptor pocket when human receptor is bound, compared to avian receptor (Extended Data Fig. 4), and in this important regard, the binding of human receptor to H10 is reminiscent of its binding to H1 HAs. Another similarity between H10 and H1 is that in both cases the residue at position 222 interacts with the 3-hydroxyl (3-OH) of Gal-2. In the case of H1, residue 222 is a lysine[18], whereas in H10 it is a glutamine (Fig. 2a); both conserved within their respective clades. In the case of H10 there is an additional interaction of the 3-OH of Gal-2 with the main-chain carbonyl at Gly 225. Thus, although the difference in orientation of Gal-2 is about 45° between the two HAs, important interactions with the 3-OH of Gal-2 are made in both cases.

The rotation of Gal-2 around its C6–C5 bond in the human receptor complex formed by H10 HA is similar to that seen in the H7 HA–human receptor complex. As detailed before[3] there is an approximately 45° rotation about the C6–C5 bond in the H7 HA–human receptor complex, relative to that seen in the complex formed between avian H7 and human receptor, that orients Gal-2 towards the 190-helix. This arrangement was correlated with the presence of Leu 226 in H7 HA. Intriguingly, however, avian H10 is able to bind human receptor in a very similar conformation to that seen in human H7 HA, even though unlike H7 it has glutamine at position 226. Thus, the substantial avidity of avian H10 for human receptor seems to be achieved by features that are reminiscent of both pandemic H1, which is phylogenetically distinct, and human

a H10 versus H1   **b** H10 versus H3   **c** H10 versus H7
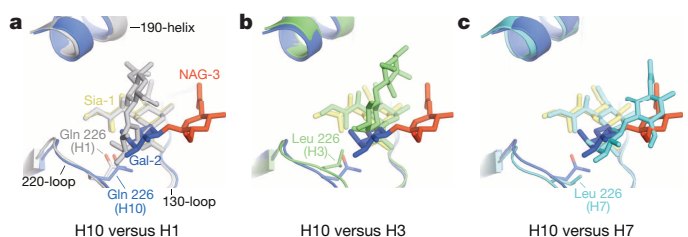
Figure 3 | Comparison of human receptor complexes of H10 with human viruses. Avian H10 overlapped with pandemic H1 (coloured grey) (**a**), pandemic H3 (coloured green) (**b**) and human H7 (coloured light blue) (**c**). The HAs from both pandemic viruses, H1 and H3, show the human receptor exiting almost vertically from the binding site, as viewed in the figure. By contrast, the human receptor bound to avian H10 exits to the side of the site as a consequence of a 45° difference in rotation about the C5–C6 bond. The Gal-2 of human receptor bound to H10 sits lower in the binding site than is the case for H3 and, in this respect, is more similar to the situation seen in H1 as shown in panels **b** and **a**, respectively. H10 shares with H1, but not H3, the characteristic of having a glutamine at position 226 (H3 numbering) and the lower positioning of the Gal-2 is correlated with the ability of the glutamine side chain to sit lower in the base of the receptor-binding site when the human receptor is bound compared to the complex with avian receptor. The higher positioning of Gal-2 in the H3 complex (and also in the other characterized pandemic H2 (ref. 5)) is correlated with the replacement of glutamine at position 226 by leucine. However, the overall trajectory with which the second and third sugars of the human receptor exit the binding site of H10 is very similar to the arrangement observed in human, but not avian[3], H7 HA (**c**).

H10–human receptor complex   H10–avian receptor complex

Figure 2 | Crystal structures of receptor complexes of H10 HA. The receptor-binding site of H10, bounded by the 130- and 220-loops and the 190-helix, is shown as a ribbons cartoon, with the receptors shown in stick representation, together with electron density from a $2F_o − F_c$ Fourier map contoured at 0.8σ. **a, b**, The human receptor complex of H10 shows the second and third sugars exiting the binding site across the 130-loop (**b**) whereas the avian receptor complex (**b**), shown overlapped with the same ligand from its complex with avian H7 HA, binds in a classical avian manner with the second and third sugars exiting the binding site over the 220-loop.

H7, which belongs to the same clade. The structure of the complex formed by human H10 HA with human receptor is very similar to that formed by avian H10 HA. It indicates in addition that Arg 137 has the potential to interact with NAG-3 of the receptor (Extended Data Fig. 1b). These interactions are consistent with affinity measurements also made using expressed human H10 HA that indicate a closer similarity between human and avian receptor binding in this case (Extended Data Fig. 1c).

The structure of the avian HA–receptor complex formed by avian H10 is like that formed by the vast majority of avian viruses[9,19] (Fig. 2b). In addition, examination of the structure of H10 HA in complex with sulphated 3′-SLN shows that the sulphate of the receptor analogue is oriented towards Lys 158A, one of the inserted residues in the protruding 150-loop (Extended Data Fig. 3a). Comparison of this structure with that of the complex formed between H5 HA and sulphated 3′-SLN[20] shows that, in the H5 complex the sulphate is oriented towards Lys 193 (Extended Data Fig. 3b). In both cases the sulphate appears to be important for the high avidities observed for this receptor.

Currently all three avian viruses considered to threaten human health—H5N1, H7N9 and H10N8—have important similarities. They are able to cause serious infections in humans but cannot be transmitted between humans; their avidity for human receptors is high, but as avian viruses they retain a preference for avian receptors. Whether or not any of these is a precursor of a pandemic virus remains to be seen. Our data suggest that in its present form H10N8 will require the sort of mutations that occurred in potential precursors of previous pandemic viruses. These mutations primarily brought about a switch from preference for avian receptor to preference for human receptor, with the accompanying gain of transmissibility between humans. In each case the switch was achieved by decreasing avidity for avian receptor. The fact that H10N8 has substantial avidity for human receptors has allowed it to overcome the initial species barrier between birds and humans. As a consequence the virus has the opportunity to adapt all of its gene products to optimize virus replication in humans. Of paramount importance in these processes of mutant selection would be those modifications in the HA function of receptor binding, the essential first step in generating a pandemic virus.

METHODS SUMMARY

A recombinant influenza virus created by reverse genetics expressing the avian A/mallard/Sweden/51/2002 H10N2 HA[17] was grown in embryonated hens' eggs. HA of recently isolated A/Jiangxi-Donghu/346/2013 (H10N8) from humans was expressed in Sf9 cells. H10 proteins were purified as described previously[21,26]. Avian H10 was crystallized in 18% PEG 2000 MME, 0.1 M bicine, pH 9.0, and cryoprotected by 25% glycerol; human H10 was crystallized in 10% PEG 6000, 5% MPD and 0.1 M HEPES, pH 7.5, and cryoprotected by 30% MPD. Crystals were soaked in cryobuffer supplemented with LSTa or 6′-SLN or sulphated 3′-SLN. Data sets were collected on beamlines IO3/IO4 (Diamond Light Source, Harwell, Oxfordshire, UK) at 100 K. Diffraction data were integrated using Denzo and scaled with Scalepack[22]. The structure was solved by molecular replacement with Phaser[23] using H7 A/turkey/Italy/214845/2002 (H7N3) (ref. 11) as a search model. Model building was performed with the program Coot[24] and structure refinement with Refmac 5 (ref. 25). Virus binding to defined receptor analogues was measured on an Octet RED biolayer interferometer (Pall ForteBio)[26]. Biotinylated sugars (3′-SLN, 6′-SLN, Le[X], sulphated Le[X] and sulphated 3′-SLN) were from Lectinity Holding and were immobilized on strepatavidin biosensors. Virus binding (100 pM virus) was measured at 25 °C in a 30–50 min association step. All solutions also contained 10 μM oseltamivir carboxylate (Roche) and 10 μM zanamivir (GSK) to prevent cleavage of the receptor analogues by the viral neuraminidase.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Chen, H. et al. Clinical and epidemiological characteristics of a fatal case of avian influenza A H10N8 virus infection: a descriptive study. Lancet (2014).
2. Gamblin, S. J. et al. The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science 303, 1838–1842 (2004).
3. Xiong, X. et al. Receptor binding by an H7N9 influenza virus from humans. Nature 499, 496–499 (2013).
4. Couceiro, J. N., Paulson, J. C. & Baum, L. G. Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. Virus Res. 29, 155–165 (1993).
5. Liu, J. et al. Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. Proc. Natl Acad. Sci. USA 106, 17175–17180 (2009).
6. Xu, R., McBride, R., Nycholat, C. M., Paulson, J. C. & Wilson, I. A. Structural characterization of the hemagglutinin receptor specificity from the 2009 H1N1 influenza pandemic. J. Virol. 86, 982–990 (2012).
7. Xu, R. & Wilson, I. A. Structural characterization of an early fusion intermediate of influenza virus hemagglutinin. J. Virol. 85, 5172–5182 (2011).
8. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. Virology 309, 209–218 (2003).
9. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. Proc. Natl Acad. Sci. USA 98, 11181–11186 (2001).
10. Zhang, W. et al. An airborne transmissible avian influenza H5 hemagglutinin seen at the atomic level. Science 340, 1463–1467 (2013).
11. Russell, R. J. et al. H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. Virology 325, 287–296 (2004).
12. Shi, Y. et al. Structures and receptor binding of hemagglutinins from human-infecting H7N9 influenza viruses. Science 342, 243–247 (2013).
13. Gambaryan, A. S. et al. 6-sulfo sialyl Lewis X is the common receptor determinant recognized by H5, H6, H7 and H9 influenza viruses of terrestrial poultry. Virol. J. 5, 85 (2008).
14. Matrosovich, M., Matrosovich, T., Uhlendorff, J., Garten, W. & Klenk, H. D. Avian-virus-like receptor specificity of the hemagglutinin impedes influenza virus replication in cultures of human airway epithelium. Virology 361, 384–390 (2007).
15. Rogers, G. N. et al. Host-mediated selection of influenza virus receptor variants. Sialic acid-α2,6Gal-specific clones of A/duck/Ukraine/1/63 revert to sialic acid-α2,3Gal-specific wild type in ovo. J. Biol. Chem. 260, 7362–7367 (1985).
16. Matrosovich, M. et al. Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. J. Virol. 74, 8502–8512 (2000).
17. Keawcharoen, J. et al. Repository of Eurasian influenza A virus hemagglutinin and neuraminidase reverse genetics vectors and recombinant viruses. Vaccine 28, 5803–5809 (2010).
18. Reid, A. H., Fanning, T. G., Hultin, J. V. & Taubenberger, J. K. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. Proc. Natl Acad. Sci. USA 96, 1651–1656 (1999).
19. Eisen, M. B., Sabesan, S., Skehel, J. J. & Wiley, D. C. Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography. Virology 232, 19–31 (1997).
20. Xiong, X. et al. Recognition of sulphated and fucosylated receptor sialosides by A/Vietnam/1194/2004 (H5N1) influenza virus. Virus Res. 178, 12–14 (2013).
21. Skehel, J. J. & Schild, G. C. The polypeptide composition of influenza A viruses. Virology 44, 396–408 (1971).
22. Otwinowski, Z. & Minor, W. Data collection and processing. Proceedings of the CCP4 Study Weekend 556–562 (SERC Daresbury Laboratory, 1993).
23. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D 63, 32–41 (2007).
24. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).
25. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240–255 (1997).
26. Xiong, X. et al. Receptor binding by a ferret-transmissible H5 avian influenza virus. Nature 497, 392–396 (2013).

Author Contributions All authors performed experiments and contributed to the writing of the manuscript.

Author Information Structural data have been deposited with the Protein Data Bank under accession numbers 4CYV, 4CYW, 4CYZ, 4CZ0 and 4D00. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.G. (sgambli@nimr.mrc.ac.uk) or J.J.S. (skehelj@nimr.mrc.ac.uk).

METHODS

Purification of HAs. A recombinant influenza virus bearing A/mallard/Sweden/51/2002 (H10N2) HA[17] was grown in embryonated hens' eggs and purified essentially as before[21] by ultracentrifugation on a sucrose gradient. H10 protein was purified from the virus membrane following bromelain digestion, Q Sepharose chromatography and finally size-exclusion chromatography on a Superdex 200 column. The ectodomain (residues 1–506 of mature HA sequence) of A/Jiangxi-Donghu/346/2013 (H10N8) was expressed in Sf9 cells after being cloned into a modified pFastBac1 encoding polyhedrin signal peptide, trimerization foldon and His-tag, essentially as before[27]. HA was purified following Co-NTA chromatography, trypsin digestion, and size-exclusion chromatography on a Superdex 200 column.

Crystallization. Avian H10 crystals were grown at 18 °C using 18% PEG 2000 MME and 0.1 M bicine, pH 9.0, and prepared for flash cooling by adding 25% glycerol. Human H10 crystals were grown at 18 °C using 10% PEG 6000, 5% MPD and 0.1 M HEPES, pH 7.5, and prepared for flash cooling by adding 30% MPD. Crystals were soaked overnight in 40 mM LSTa, 3′-SLN, 6′-SLN or sulphated 3′-SLN in cryobuffer. Data sets were collected at Diamond Light Source, Harwell, Oxfordshire, UK at 100 K. Diffraction data were integrated using XDS and Scala. The structures were solved by molecular replacement with Phaser[23] using H7 A/Turkey/Italy/02/[11] as a search model. Model building was performed with the program Coot[24] and structure refinement with Refmac 5 (ref. 25).

BLI. Virus binding to defined receptor analogues was measured on an Octet RED biolayer interferometer (Pall ForteBio) essentially as described before[26]. Biotinylated polymeric sugars (3′-SLN, 6′-SLN, Le[X], sulphated Le[X] and sulphated 3′-SLN) containing 20% mol. sugar, 5% mol. biotin on a 30 kDa polyacrylamide backbone were from Lectinity Holding and were immobilized on streptavidin biosensors at concentrations ranging from 0.01 to 1.5 µg ml^{-1}. Virus binding (at 100 pM virus) was measured at 25 °C in a 30–50 min association step. All solutions also contained 10 µM oseltamivir carboxylate (Roche) and 10 µM zanamivir (GSK) to prevent cleavage of the receptor analogues by the viral neuraminidase.

MST. MST measurements were performed using a NanoTemper Monolith NT.115 instrument (NanoTemper Technologies GmbH) following essentially the same protocol as before[26]. HA samples were labelled with the amine-reactive dye NT-647 using the Monolith NT.115 Protein Labelling Kit RED-NHS. Sodium salts of 3′-SLN and 6′-SLN (Dextra) were dissolved to a final concentration of 50–100 mM in PBS buffer containing 0.05% Tween-20 and labelled HA at a concentration of ∼80 nM. This stock solution was then serially diluted 1:1 using the same buffer to give ten working solutions with different sugar concentrations but the same fluorophore concentration. These solutions were then loaded into hydrophilic treated capillaries and MST measurements were made at 25 °C using 30% light-emitting diode power and 45% infrared-laser power. The laser-on time was 30 s and laser-off time 5 s. All measurements were made at least five times.

27. Lin, Y. P. *et al.* Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin. *Proc. Natl Acad. Sci. USA* **109,** 21474–21479 (2012).
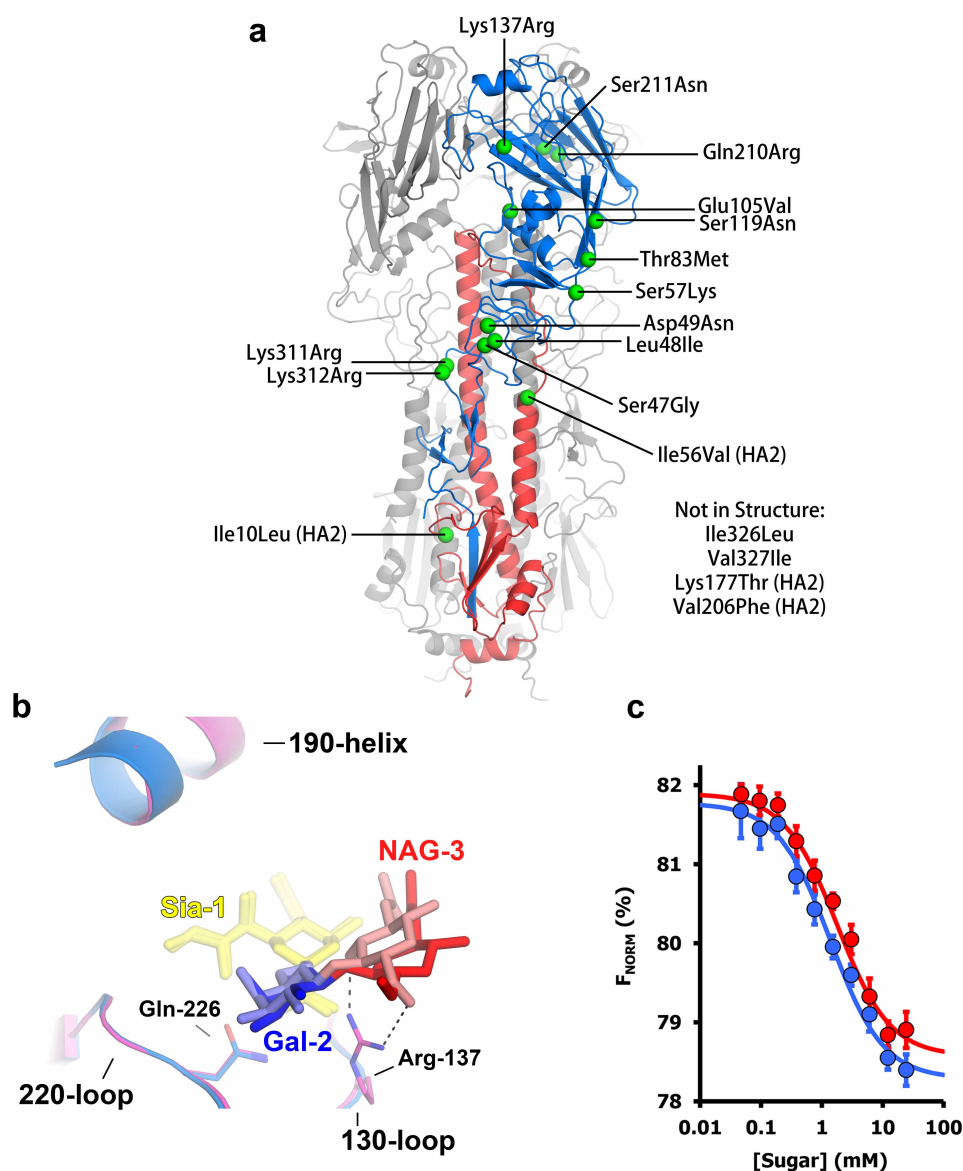
Extended Data Figure 1 | The structures of avian H10 HA and human H10 HA. **a,** The structure of avian H10 HA trimer with one subunit coloured blue (HA1) and one coloured red (HA2), and the other two coloured grey. The locations of amino acid sequence differences between the HAs of A/mallard/Sweden/51/2002(H10N2) and A/Jiangxi-Donghu/346/2013 (H10N8) from humans are indicated by green spheres and numbered by alignment with H3 sequences. **b,** Receptor-binding site of human H10 HA (purple) compared to that of avian H10 HA (blue). The structure of the human H10 has been determined to 2.5 Å. Human receptors bound to avian H10 are coloured yellow (Sia-1), blue (Gal-2) and red (NAG-3), whereas the equivalents from the human H10 complex are shown in lighter shades. Potential hydrogen bonds between Arg 137 of human H10 and the human receptor are indicated by dashed lines. The arginine residue potentially also makes other hydrogen bonds (not shown) including to the glycosidic oxygen. The crystallographic asymmetric unit contains one HA trimer, the figure shows the A-chain monomer which was selected on the basis of not being involved in a close crystal contact and for having well-ordered electron density for the arginine. **c,** Binding of NT647-labelled human H10 to 3′-SLN (red symbols) and 6′-SLN (blue symbols). The calculated K_ds were 1.81 ± 0.39 mM and 1.39 ± 0.32 mM, respectively.
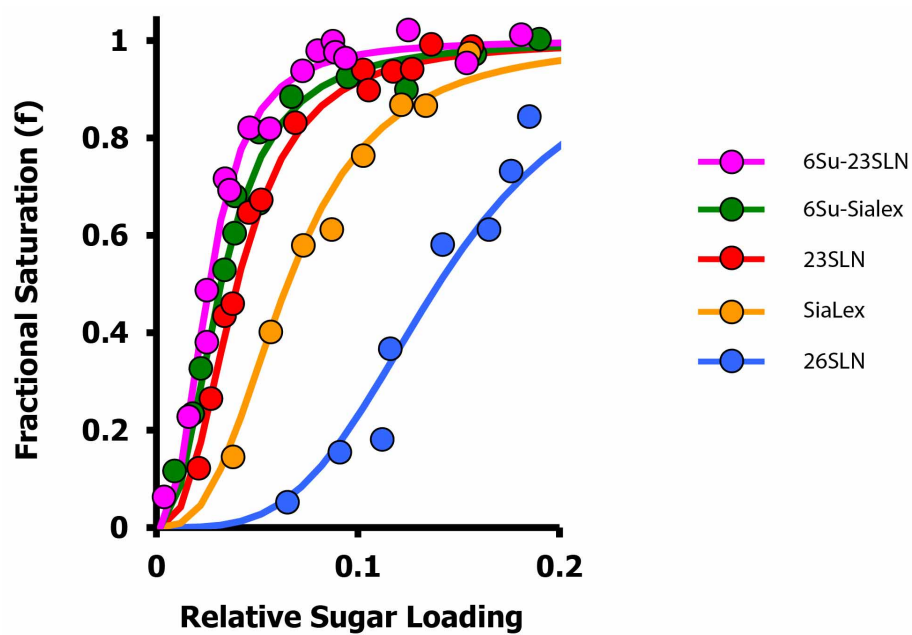
Extended Data Figure 2 | Biolayer interferometry measurements of binding avidity. Biolayer interferometry binding data for A/mallard/Sweden/51/2002 (H10N2) virus (ref. 17) to sulphated 3′-SLN (purple), sulphated SLe^X (green), 3′-SLN (red), SLe^X (orange) and 6′-SLN (blue).

H10 Sulfated 3SLN

H5 Sulfated 3SLN
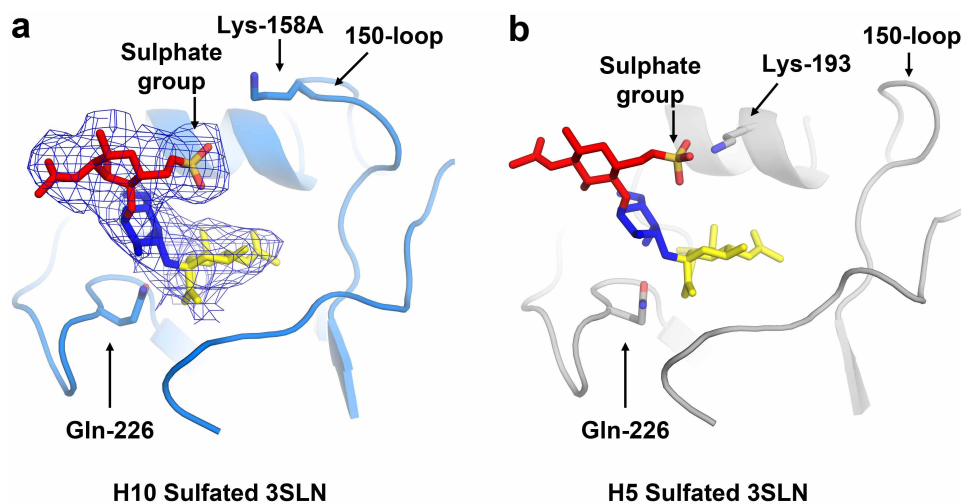
Extended Data Figure 3 | Structural comparison of the H10 and H5 HA binding sites. A comparison of the receptor-binding sites of H10 (**a**) and H5 (**b**) (ref. 20) HAs from complexes formed between the HAs and sulphated 3′-SLN. Electron density ($2F_c – F_o$, 0.8σ) for the receptor is shown in **a**. In H10 HA the sulphate group approaches Lys 158A, the first inserted residue in the 150-loop. By contrast, in H5 HA the sulphate approaches Lys 193 in the 190-helix (**b**) (the equivalent residue in H10 HA is aspartic acid).

a H10 avian and human receptor bound forms compared

b H1 avian and human receptor bound forms compared
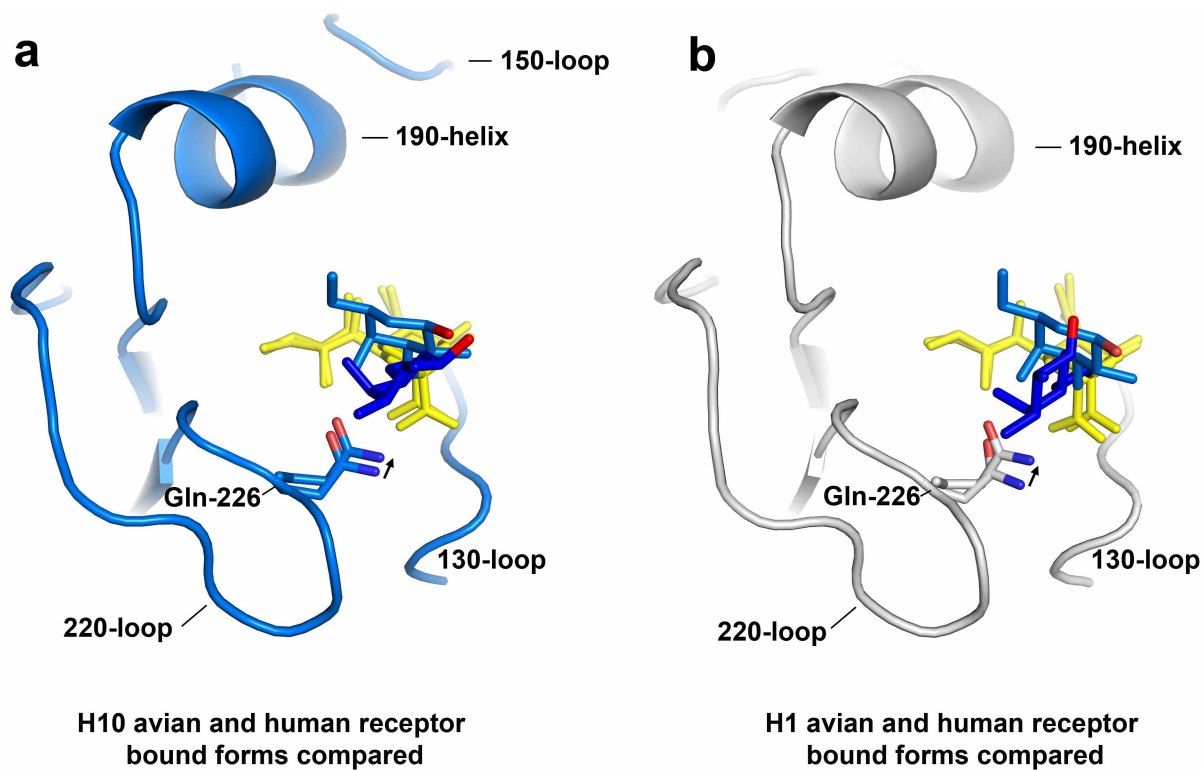
Extended Data Figure 4 | Comparison of H10 and H1 HA in complex with avian and human receptors. Comparisons of avian- and human-receptor-bound forms of H10 (**a**) and H1 (**b**) HAs. The arrows indicate the upward movement of Gln 226 in the complexes formed with avian receptor.

Extended Data Table 1 | Table of crystallographic statistics for the structure determined

Haemagglutinin	Avian H10				Human H10
Receptor Analogue	APO	α2,6SLN	LSTA	6-O-sulfated-2,3SLN	α2,6SLN
Data collection					
X-ray Source	DIAMOND I02	DIAMOND I04	DIAMOND I04	DIAMOND I04-1	Rigaku MicroMax-007 HF
Wavelength (Å)	0.92	0.92	0.98	0.92	1.54
Resolution range (Å)	49.15 - 2.3 (2.34 - 2.3)	48.85 - 2.60 (2.66 - 2.60)	62.86 -2.4 (2.53 - 2.4)	48.93 - 3.2 (3.37 - 3.2)	40 - 2.5 (2.61 - 2.5)
Space group	P21	P21	P21	P21	P2$_1$2$_1$2$_1$
a, b, c (Å)	67.61 227.99 68.72	68.97, 215.44, 79.36	67.07, 230.58, 68.46	69.11, 216.09, 81.49	90.601, 183.17, 192.15
α, β, γ (°)	90, 110.32 , 90	90, 104.91, 90	90, 110.41, 90	90, 105.22, 90	90, 90, 90
Total reflections	299791	242096	296672	132235	390165
Unique reflections	84779	67905	75050	37572	103064
Multiplicity	3.5 (3.5)	3.6 (3.7)	4.0 (4.0)	3.5 (3.6)	3.8 (3.2)
Completeness (%)	98.4 (99.0)	99.1 (99.8)	99.1 (99.2)	98.9 (99.5)	92.8 (90.0)
Mean I/sigma(I)	11.9 (2.3)	11.8 (1.8)	9.5 (3.0)	7.2 (1.4)	6.21 (2.06)
R-sym	0.061 (0.52)	0.073 (0.691)	0.110 (0.703)	0.174 (0.946)	0.163 (0.518)
Refinement					
R-factor	0.2138	0.2082	0.2045	0.2903	0.2085
R-free	0.2520	0.2536	0.2442	0.308	0.2347
Number of atoms	12135	12105	12251	11837	12576
macromolecules	11424	11427	11713	11687	11667
receptor analogue	-	137	77	150	138
solvent	380	285	421	0	11529
Protein residues	1472	1475	1470	1470	1483
RMSD Bond Length (Å)	0.0121	0.005	0.0069	0.0055	0.006
RMSD Bond Angle (°)	1.606	1.000	1.09	1.0784	1.1023
Ramachandran favored (%)	95.46	96.79	98.22	96.78	98.16
Ramachandran outliers (%)	0.0	0.14	0	0.14	0
Average B-factor (Å^2)	57.075	62.09	47.80	96.43	30.99
macromolecules	57.00	62.48	46.30	96.11	31.19
receptor analogue	-	101.45	53.20	121.38	63.60
solvent	43.70	45.89	37.40	-	28.31
PDB code	4CYV	4CYW	4CYZ	4CZ0	4D00

LETTER

Ultraviolet radiation accelerates BRAF–driven melanomagenesis by targeting TP53

Amaya Viros[1]*, Berta Sanchez-Laorden[1]*, Malin Pedersen[2]*, Simon J. Furney[1]*, Joel Rae[2], Kate Hogan[1], Sarah Ejiama[1], Maria Romina Girotti[1], Martin Cook[1,3], Nathalie Dhomen[2] & Richard Marais[1,2]

Cutaneous melanoma is epidemiologically linked to ultraviolet radiation (UVR), but the molecular mechanisms by which UVR drives melanomagenesis remain unclear[1,2]. The most common somatic mutation in melanoma is a V600E substitution in BRAF, which is an early event[3]. To investigate how UVR accelerates oncogenic BRAF-driven melanomagenesis, we used a BRAF(V600E) mouse model. In mice expressing BRAF(V600E) in their melanocytes, a single dose of UVR that mimicked mild sunburn in humans induced clonal expansion of the melanocytes, and repeated doses of UVR increased melanoma burden. Here we show that sunscreen (UVA superior, UVB sun protection factor (SPF) 50) delayed the onset of UVR-driven melanoma, but only provided partial protection. The UVR-exposed tumours showed increased numbers of single nucleotide variants and we observed mutations (H39Y, S124F, R245C, R270C, C272G) in the *Trp53* tumour suppressor in approximately 40% of cases. *TP53* is an accepted UVR target in human non-melanoma skin cancer, but is not thought to have a major role in melanoma[4]. However, we show that, in mice, mutant *Trp53* accelerated BRAF(V600E)-driven melanomagenesis, and that *TP53* mutations are linked to evidence of UVR-induced DNA damage in human melanoma. Thus, we provide mechanistic insight into epidemiological data linking UVR to acquired naevi in humans[5]. Furthermore, we identify *TP53/Trp53* as a UVR-target gene that cooperates with BRAF(V600E) to induce melanoma, providing molecular insight into how UVR accelerates melanomagenesis. Our study validates public health campaigns that promote sunscreen protection for individuals at risk of melanoma.

We expressed BRAF(V600E) in the melanocytes of 2-month-old mice[6] and, 1 month later, protected half of the shaved backs of the mice with a cloth and exposed the other half to low doses of UVR (Fig. 1a and Extended Data Fig. 1a, b). UVR upregulates Trp53 (ref. 7), and, within 24 h of a single UVR exposure, we observed Trp53 staining in ~41% of epidermal keratinocytes and down into the reticular dermis (Fig. 1b, Extended Data Fig. 1c and Extended Data Table 1). We also observed abundant sunburn cells (~17 cells mm^{-1}) in the basal layer of the UVR-exposed skin (Fig. 1c and Extended Data Table 1). The skin did not blister, but after 24–48 h it developed mild erythema, and after 7 days it was rough to the touch, showed fine desquamation and was thicker due to hyperkeratosis, epidermal hypertrophy and a thickened fibrotic dermis (Fig. 1d).

UVR induces melanocyte proliferation[8], and naevogenesis is driven by clonal expansion of melanocytes expressing mutant *BRAF*, the acquisition of which is an early event[3]. Furthermore, the incidence of acquired human naevi, ~80% of which carry *BRAF* mutations[9], is linked to UVR exposure[10]. We show that UVR induced BRAF(V600E)-melanocyte proliferation *in vivo* in mice (Extended Data Fig. 2) and, within 7 days, the UVR-exposed skin had more abundant and larger naevi than non-UVR-exposed skin (Fig. 1d, Extended Data Fig. 3a–d and Extended Data Table 2). These UVR-induced changes persisted (Fig. 1e) and the UVR-exposed skin darkened (Fig. 1f). UVR did not induce naevi or skin darkening in

non-BRAF(V600E) mice (Extended Data Fig. 3e, f). Thus, we validate human epidemiological studies by showing that BRAF(V600E)-expressing melanocytes are susceptible to proliferation and naevogenesis driven by low-dose UVR that mimics mild sunburn.

As previously reported[6], BRAF(V600E) induced melanoma in ~70% of mice at a median latency of 12.6 months. On average each mouse developed 0.9 tumours on its back (Fig. 2a, b). When exposed to UVR, all BRAF(V600E) mice developed melanoma within 7 months at a median latency of 5.3 months and an average of ~3.5 (range 1–6) tumours each, 98% (59/60) of which were within the UVR-exposed area (Fig. 2a–c). UVR did not induce melanoma in non-BRAF(V600E) mice (Fig. 2a). Thus, we concluded that UVR accelerated BRAF(V600E)-driven melanomagenesis.

The UVR and non-UVR tumours were asymmetrical, predominantly amelanotic and locally destructive (Fig. 2d and Extended Data Fig. 4a, b). They invaded the deeper subcutis, showed superficial naevi and pigmented cells, and ~50% ulcerated the overlying epidermis (Fig. 2e and Extended Data Fig. 4b). The tumour cells were atypical with dendritic/plump spindle morphology, and stained positive for S100 and HMB-45/MelanA (Fig. 2e, f and Extended Data Fig. 4a, c–e). They were heterogeneous, highly mitotic (20 mitoses per 10 high-power fields; 10 tumours analysed) and Ki-67 positive (26.4%, range 10–50; Extended Data Fig. 4a, f).

Application of sunscreen (UVA superior, UVB SPF 50) 30 min before UVR exposure blocked induction of Trp53 and apoptosis in epidermal keratinocytes (Extended Data Fig. 5a–c and Extended Data Table 1). These are largely UVB-driven responses[7], but sunscreen also blocked UVA-induced epidermal hypertrophy and dermal thickening (Extended Data Fig. 5d). The cloth-protected and sunscreen-protected naevi were indistinguishable and sunscreen prevented UVR-driven skin darkening (Extended Data Fig. 6a, b). Thus, sunscreen protected the epidermis from the immediate effects of UVA and UVB, but nevertheless all of the sunscreen-protected UVR-exposed BRAF(V600E) mice developed tumours within 15 months at a median latency of 7.5 months (Fig. 2a), a significant reduction compared with unprotected mice, but a significant increase compared with non-UVR-exposed mice ($P = 0.03$; Fisher's exact test). Sunscreen-protected mice developed ~1.5 tumours on their backs, again significantly fewer than unprotected mice, but significantly more than non-UVR-exposed mice (Fig. 2b, c). The sunscreen-protected melanomas had similar histology to non-UVR and UVR-exposed melanomas (Fig. 2d–f and Extended Data Fig. 6c) and were more common on sunscreen-protected (14 tumours) than cloth-protected (8 tumours) skin. Thus, sunscreen only partially prevented acceleration of BRAF(V600E)-driven melanomagenesis by UVR.

Array comparative genome hybridization (aCGH) revealed no significant differences in gross chromosomal aberrations or copy number variations between non-UVR and UVR melanomas (Fig. 3a), but the non-UVR melanomas showed significantly more (60% versus 13.3%) recurrent copy number changes in melanoma oncogenes and tumour suppressors[11] (*Mitf, Braf, Ccnd1, Cdkn2a*; Fig. 3b and Supplementary Table 1). Whole-exome sequencing (WES) revealed that UVR-exposed

[1]Molecular Oncology Group, Cancer Research UK Manchester Institute, University of Manchester, Wilmslow Road, Manchester M20 4BX, UK. [2]Signal Transduction Team, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK. [3]Histopathology, Royal Surrey County Hospital, Egerton Road, Guildford GU2 7XX, UK.
*These authors contributed equally to this work.
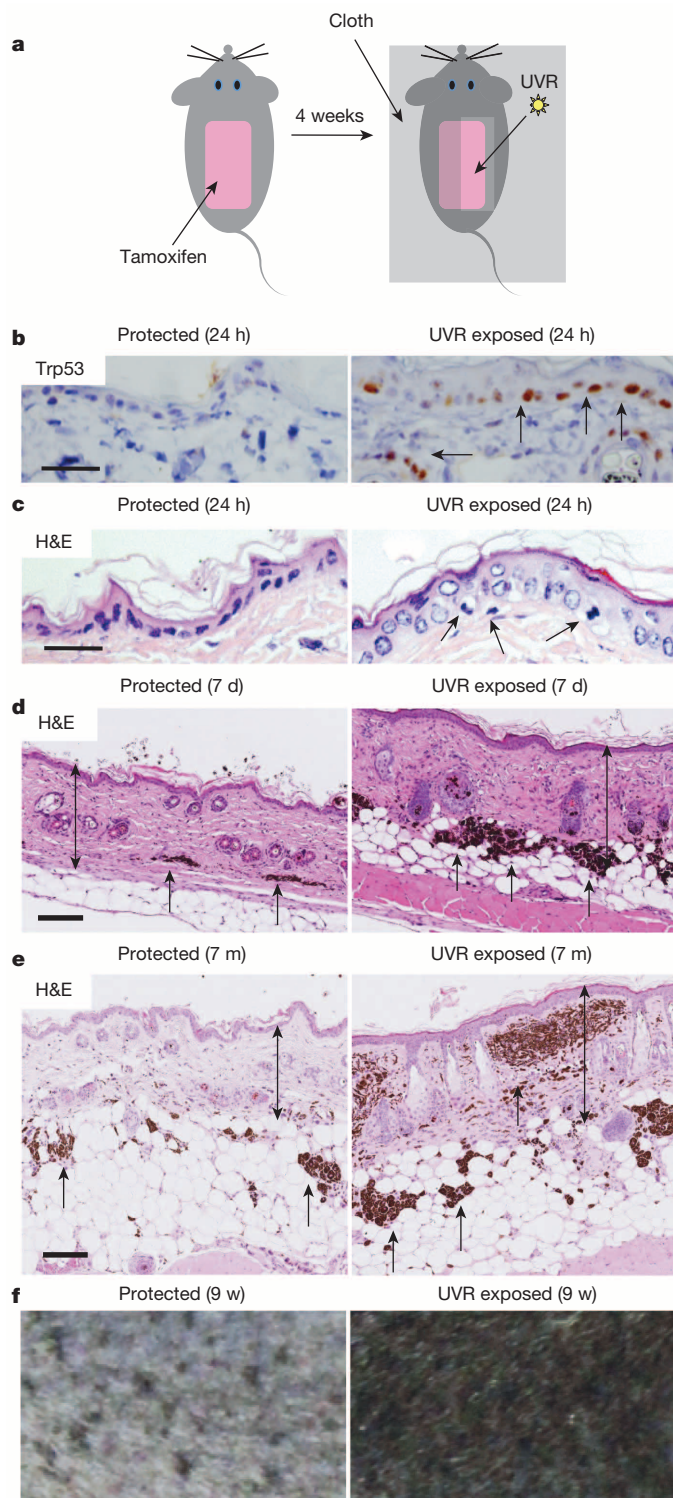
Figure 1 | UVR accelerates BRAF(V600E)-driven naevogenesis.
a, Experimental design. **b**, Trp53 staining (arrows) in protected and UVR-exposed epidermis 24 h after UVR. Scale bar, 50 μm. Five animals were examined. **c**, Haematoxylin and eosin (H&E)-stained section of protected and UVR-exposed epidermis 24 h after UVR. Arrows indicate sunburn cells (apoptotic keratinocytes). Scale bar, 50 μm. **d**, H&E stain of protected and UVR-exposed epidermis 7 days (d) after UVR. Arrows indicate dermal naevi; double-headed arrows indicate epidermal–dermal thickness. Scale bar, 500 μm. **e**, H&E stain of protected and UVR-exposed epidermis 7 months (m) after UVR. Arrows indicate dermal naevi; double-headed arrows indicate epidermal–dermal thickness. Scale bar, 500 μm. **f**, Protected and UVR-exposed skin 9 weeks (w) after UVR treatments. Original magnification, ×2.
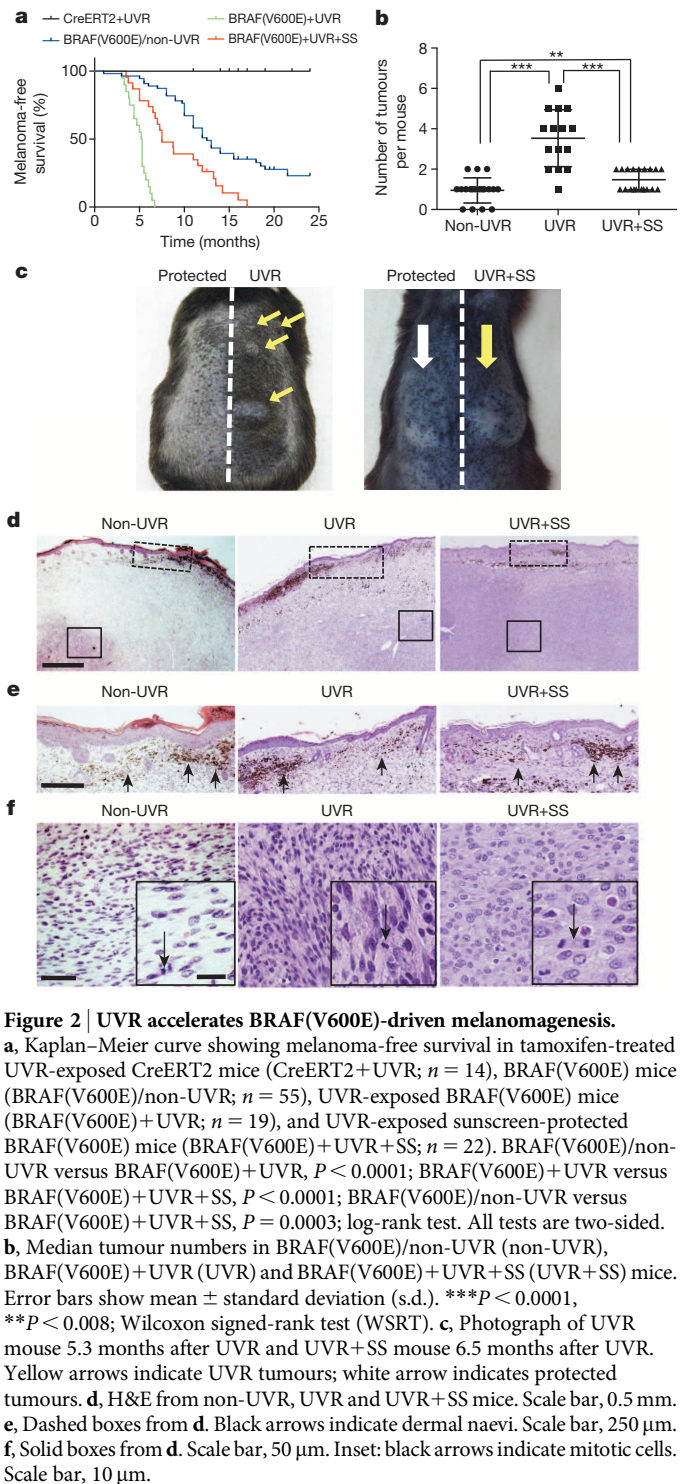
Figure 2 | UVR accelerates BRAF(V600E)-driven melanomagenesis.
a, Kaplan–Meier curve showing melanoma-free survival in tamoxifen-treated UVR-exposed CreERT2 mice (CreERT2+UVR; $n = 14$), BRAF(V600E) mice (BRAF(V600E)/non-UVR; $n = 55$), UVR-exposed BRAF(V600E) mice (BRAF(V600E)+UVR; $n = 19$), and UVR-exposed sunscreen-protected BRAF(V600E) mice (BRAF(V600E)+UVR+SS; $n = 22$). BRAF(V600E)/non-UVR versus BRAF(V600E)+UVR, $P < 0.0001$; BRAF(V600E)+UVR versus BRAF(V600E)+UVR+SS, $P < 0.0001$; BRAF(V600E)/non-UVR versus BRAF(V600E)+UVR+SS, $P = 0.0003$; log-rank test. All tests are two-sided.
b, Median tumour numbers in BRAF(V600E)/non-UVR (non-UVR), BRAF(V600E)+UVR (UVR) and BRAF(V600E)+UVR+SS (UVR+SS) mice. Error bars show mean ± standard deviation (s.d.). ***$P < 0.0001$, **$P < 0.008$; Wilcoxon signed-rank test (WSRT). **c**, Photograph of UVR mouse 5.3 months after UVR and UVR+SS mouse 6.5 months after UVR. Yellow arrows indicate UVR tumours; white arrow indicates protected tumours. **d**, H&E from non-UVR, UVR and UVR+SS mice. Scale bar, 0.5 mm. **e**, Dashed boxes from **d**. Black arrows indicate dermal naevi. Scale bar, 250 μm. **f**, Solid boxes from **d**. Scale bar, 50 μm. Inset: black arrows indicate mitotic cells. Scale bar, 10 μm.

melanomas had significantly more somatic single nucleotide variants (SNVs; median 6.16 (range 0.26–40.69) versus 0.19 (range 0.02–2.16) SNVs per megabase pair (Mb); Fig. 3c and Supplementary Table 1) and a significantly higher proportion of C-to-T transitions (median 16.7% versus 76.6%) at the 3′ end of pyrimidine dimers than non-UVR tumours (Fig. 3d, e). This provides direct evidence of UVR-induced DNA damage[12] and, notably, although sunscreen significantly reduced the number of SNVs (median 0.25050 SNVs per Mb (range 0.08233–0.64432)) in UVR-exposed melanomas (Fig. 3c and Supplementary Table 1), the tumours still showed a higher proportion of C-to-T transitions at the 3′ end of pyrimidine dimers than non-UVR-exposed melanomas (Fig. 3d, e). Thus,
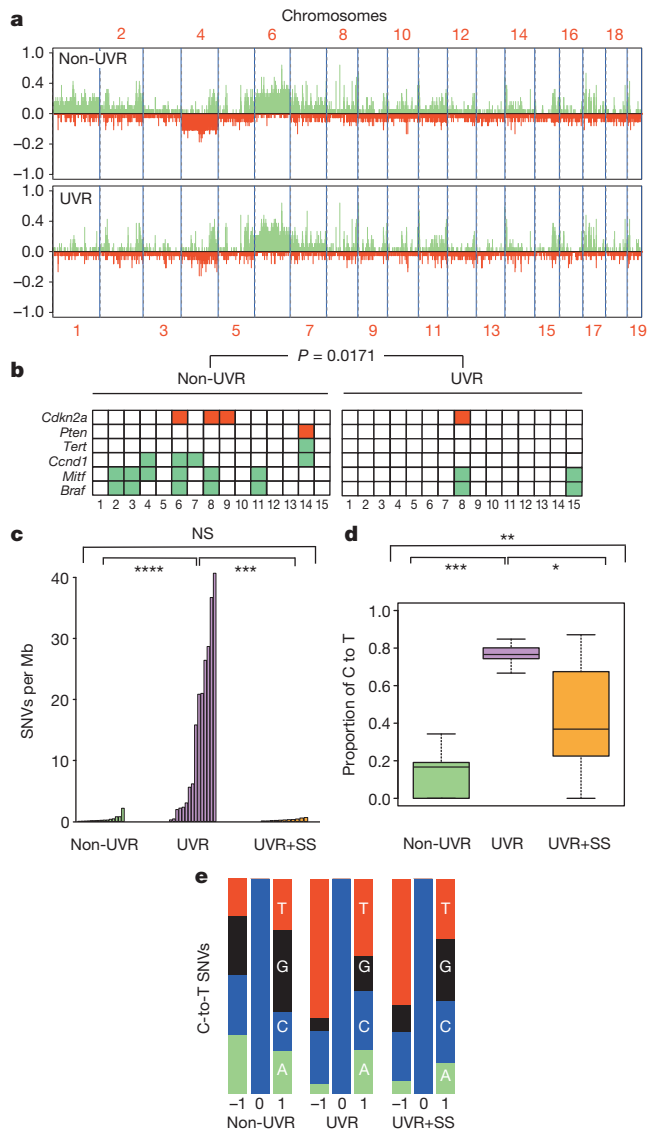
Figure 3 | Genome analysis of UVR-driven melanomas. a, Somatic copy number alterations as determined by aCGH in non-UVR and UVR tumours. *x*-axis indicates genomic position beginning at 1p and ending at 19q. *y*-axis represents the fraction of the samples with a given region gained or lost. **b**, Melanoma oncogene and tumour suppressor gene gains (green) and losses (red) as determined by aCGH in non-UVR and UVR tumours. $P = 0.0171$; WRST. **c**, Somatic SNVs per Mb in non-UVR (green), UVR (violet) and UVR+SS (orange) tumours. $****P = 1.79 \times 10^{-6}$, $***P = 1.165 \times 10^{-5}$; NS, not significant ($P = 0.6461$); WRST. **d**, Proportion of C-to-T (G-to-A) transitions at the 3′ end of pyrimidine dimers in non-UVR, UVR and UVR+SS tumours. Error bars show the lowest data still within 1.5 times the interquartile range (IQR) of the lowest quartile and highest data within 1.5 times the IQR of the upper quartile. $***P < 0.0001$, $**P = 0.007698$, $*P = 0.02558$; WRST. **e**, Proportion of each nucleotide (A: green; C: blue; G: black; T: red) ± 1 bp of all C-to-T (G-to-A) transitions in non-UVR, UVR and UVR+SS tumours.

sunscreen reduced overall mutation burden but did not completely block UVR-induced DNA damage.

We observed *Trp53* mutations in 6 of 15 UVR-exposed melanomas, 1 of 11 sunscreen-protected UVR-exposed melanomas, but none of 15 non-UVR-exposed tumours (Fig. 4a, Extended Data Table 3a and Supplementary Table 1). We confirmed the *Trp53* mutations (Fig. 4b and Extended Data Table 3a) and note that there were no significant differences in the number of SNVs in the UVR-exposed *Trp53*-mutant and *Trp53* wild-type tumours (Extended Data Fig. 7a), suggesting that mutant *Trp53* did not contribute to overall mutation burden. The mutations

involved codons H39, S124, R245, R270 and C272, and six of seven were C-to-T transitions at the 3′ end of pyrimidine dimers (Extended Data Table 3a), supporting a direct role for UVR in the induction of *Trp53* mutations in melanoma. S124, R245, R270 and C272 are in the Trp53 DNA-binding domain and the corresponding residues (S127, R248, R273 and C275) are mutated in human TP53 in melanoma (Extended Data Table 3a), producing deleterious effects that severely impair promoter-specific TP53 transcriptional activity[13] (Extended Data Table 3b). R273 and R248 are human cancer mutation hotspots and produce proteins with dominant-negative and other gain-of-function activities[14–16]. Mutations in the amino-terminus of TP53 are rare, and H39 of mouse Trp53 is not conserved in human TP53, but the equivalent residue (Q38) is mutated in human cancer (Extended Data Table 3a). Notably, all of the mutant proteins accumulated in our mouse melanomas (Fig. 4c and Extended Data Table 3a), suggesting common functional consequences.

Two conditional-inducible mouse alleles for DNA-binding-domain Trp53 mutants are available, $Trp53^{LSL-R172H}$ and $Trp53^{LSL-R270H}$. The resultant mutant proteins are functionally similar; they possess dominant-negative and gain-of-function activities, and they accumulate in tumours[16]. The $Trp53^{LSL-R172H}$ mice were readily available, so we generated heterozygous mice that expressed one copy of Trp53(R172H) in their melanocytes ($Trp53^{+/LSL-R172H};Tyr::CreERT2^{+/o}$). These mice did not develop melanoma, but when crossed to the BRAF(V600E) mice ($Trp53^{+/LSL-R172H}::Braf^{+/LSL-V600E};Tyr::CreERT2^{+/o}$), all of the mice developed on average 6 tumours (range 2–10) on their backs within 3.5 months (Fig. 4d–f). The tumours presented the histopathological features of other BRAF(V600E)-driven melanomas (Fig. 4g, h and Extended Data Fig. 7b). Thus, as with the deletion of one copy of the tumour suppressor *Pten* (Extended Data Fig. 7c)[17], we show that mutant *Trp53* accelerated BRAF-driven melanomagenesis ($P < 0.0001$; Fisher's exact test).

Finally, analysis of WES data from the Broad Institute[18], The Cancer Genome Atlas (TCGA; https://tcga-data.nci.nih.gov/tcga/) and Yale University[19] revealed that *TP53* mutations coincided with a higher proportion of C-to-T transitions and a UVR signature in primary and metastatic human cutaneous melanomas (Fig. 4i, Extended Data Fig. 7d and Extended Data Table 3c). Thus, in human and mouse melanomas, *TP53/Trp53* mutations were associated with UVR-induced DNA damage, establishing that UVR accelerates melanomagenesis by targeting *TP53/Trp53*.

TP53 is a recognized UVR target in non-melanoma skin cancers, but is thought not to have a role in melanoma, because frequent p14^ARF loss stabilizes HDM2, inactivating TP53 (ref. 4). However, we show that ~40% of mouse and ~20% of human melanomas with evidence of UVR-induced DNA damage carried *Trp53/TP53* mutations[18]. *Trp53* gene deletions can cooperate with BRAF(V600E) to induce melanoma in mice and fish[20,21], but in UVR-exposed melanomas *TP53/Trp53* is mutated rather than lost, generating proteins with gain-of-function and dominant-negative activities[14–16]. This suggests a complex role for TP53/Trp53 in melanoma that cannot be accomplished if the protein is lost. Our data establish that *TP53* is a bona fide UVR target in melanoma, providing insight into how UVR drives melanoma, albeit through mechanisms that have not yet been elucidated. Mutant TP53 drives tumorigenesis through processes inducing evasion of senescence, activation of autophagy, engagement of DNA-damage responses and oxidative stress pathway reactivation, but individual mutants influence these functions differently. Also, it was recently shown that UVR stimulates melanoma invasion, but the role of TP53 in this remains unknown[22]. Studies are required to elucidate the role of TP53 in melanoma and to identify other UVR targets in *TP53/Trp53*-mutant and *TP53/Trp53* wild-type tumours. Intriguingly, recurrent UVR-signature mutations occur in *RAC1*, *STK19* and *PPP6C* in melanoma, but their roles in melanomagenesis at present are unclear[18,19]. It is also unclear what induces BRAF (V600E) because it is not a UVR-signature mutation. Nevertheless, our data show that BRAF(V600E)-expressing melanocytes are susceptible to UVR-driven naevogenesis and melanomagenesis.

In previous UVR melanoma models, embryonic mice constitutively expressed oncogenes, generally with loss of tumour suppressor genes, and
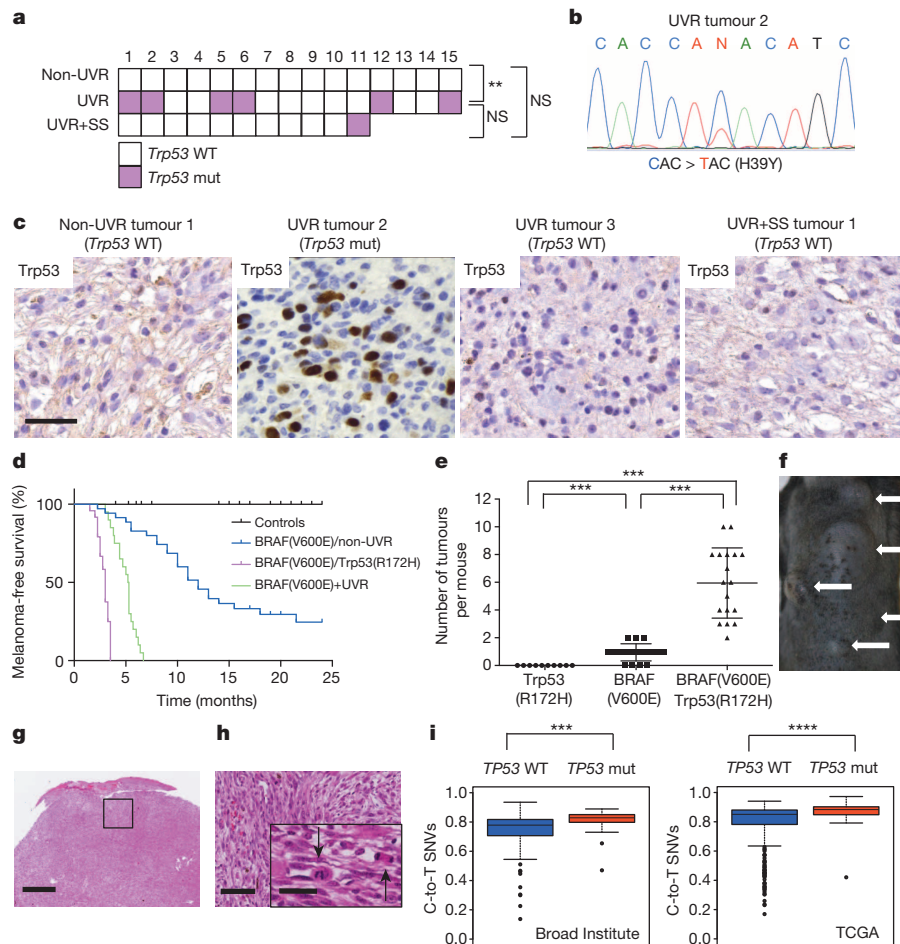
Figure 4 | Mutant *Trp53* accelerates BRAF(V600E)-driven melanomagenesis. a, *Trp53* mutations in non-UVR, UVR and UVR+SS tumours. WT, wild type (white); mut, mutant (magenta). **$P = 0.017$, Fisher's exact test. NS, not significant. **b,** Sanger sequencing confirming C-to-T, p.H39Y *Trp53* mutation in UVR tumour 2. **c,** Trp53 immunostaining in non-UVR tumour 1 (*Trp53* wild type), UVR tumour 2 (*Trp53* mutant), UVR tumour 3 (*Trp53* wild type) and UVR+SS tumour 1 (*Trp53* wild type). Scale bar, 50 μm. **d,** Kaplan–Meier curve showing melanoma-free survival in tamoxifen-treated CreERT2 and CreERT2/Trp53(R172H) (control; $n = 42$), BRAF(V600E)/non-UVR ($n = 35$), BRAF(V600E)/Trp53(R172H) ($n = 23$) and BRAF(V600E)+UVR ($n = 19$) mice. BRAF(V600E)/non-UVR versus BRAF(V600E)/Trp53(R172H), $P < 0.0001$; BRAF(V600E)+UVR versus BRAF(V600E)/Trp53(R172H), $P < 0.0001$; log-rank test. **e,** Median tumour numbers in tamoxifen-treated CreERT2/Trp53(R172H) mice (Trp53(R172H)), BRAF(V600E)/non-UVR (BRAF(V600E)) and BRAF(V600E)/Trp53(R172H) mice. Error bars show mean ± s.d. ***$P < 0.0001$; WRST. **f,** BRAF(V600E)/Trp53(R172H) mouse with multiple tumours (white arrows). **g,** H&E from a BRAF(V600E)/ Trp53(R172H) tumour. Scale bar, 0.5 mm. **h,** Boxed area from **g**. Scale bar, 250 μm. Inset: black arrows indicate mitotic cells. Scale bar, 5 μm. **i,** Proportion of C-to-T (G-to-A) transitions in TP53 wild-type (WT) or mutant (mut) human melanomas from the Broad Institute[18] (***$P = 0.002$; WRST) and TCGA (https://tcga-data.nci.nih.gov/tcga/; ****$P = 1.79 \times 10^{-5}$) data sets. Error bars show the lowest data still within 1.5 IQR of the lowest quartile and highest data within 1.5 IQR of the upper quartile.

the mice received a single burning dose of UVR 1–3 days after birth[23,24]. In this model melanoma is driven by inflammation mediated by interferon-γ secreted by neonatal macrophages, and it does not accelerate BRAF (V600E)-driven melanoma[25]. We expressed BRAF(V600E) at physiological levels in adolescent mice without manipulation of tumour suppressor genes and exposed the mice to repeated low doses of UVR. Thus, we mimicked both somatic mutation acquisition and mild sunburn in humans, and our tumours were driven by acquired *Trp53* mutations. We therefore reveal two UVR melanoma pathways, one driven by inflammation in neonates and one driven by UVR-induced mutations in adults.

It is accepted that sunscreen protects against squamous cell carcinoma[26], but controversy surrounds its ability to protect against melanoma[27–29]. The US Environmental Protection Agency states that there is "no evidence that sunscreens protect you from malignant melanoma" (http://www.epa.gov/sunwise/doc/sunscreen.pdf), but we establish that sunscreen delayed UVR-driven melanoma in susceptible mice, validating public health campaigns that promote its use for melanoma. However,

sunscreen did not safeguard against UVR completely and we advocate combining it with other sun avoidance strategies, particularly in at-risk individuals with BRAF-mutant naevi[1,30].

METHODS SUMMARY

BRAF(V600E) expression and UVR treatments. BRAF(V600E) was expressed in mouse melanocytes by topical application of tamoxifen to the shaved backs of mice at ~2 months of age as described previously[6] (Fig. 1a and Extended Data Fig. 1a). One month later, half the back of anaesthetized mice was vertically covered with a UVR-proof cloth, providing an internal contralateral control, and the other half was exposed to 160 mJ cm^{-2} UVA/UVB using a broad-spectrum UVA/UVB lamp, performing weekly re-exposures for up to 6 months (Fig. 1a and Extended Data Fig. 1a, 1b). For the sunscreen protection experiments, mice were treated with Sunsense Milk Sunscreen SPF 50 as recommended by the US Food and Drug Administration (2.2 mg cm^{-2}) by topical application 30 min before UVR exposure.
Tumour analysis. Mouse genomic DNA was prepared for aCGH and WES as described in Methods. One-hundred and seven randomly selected SNVs were validated by Sanger sequencing to reveal a true positive rate of 93.5%.

1. Whiteman, D. C. *et al.* Melanocytic nevi, solar keratoses, and divergent pathways to cutaneous melanoma. *J. Natl. Cancer Inst.* **95,** 806–812 (2003).
2. Gilchrest, B. A., Eller, M. S., Geller, A. C. & Yaar, M. The pathogenesis of melanoma induced by ultraviolet radiation. *N. Engl. J. Med.* **340,** 1341–1348 (1999).
3. Yeh, I., von Deimling, A. & Bastian, B. C. Clonal *BRAF* mutations in melanocytic nevi and initiating role of *BRAF* in melanocytic neoplasia. *J. Natl. Cancer Inst.* **105,** 917–919 (2013).
4. Tsao, H., Chin, L., Garraway, L. A. & Fisher, D. E. Melanoma: from mutations to medicine. *Genes Dev.* **26,** 1131–1155 (2012).
5. Harrison, S. L., MacLennan, R., Speare, R. & Wronski, I. Sun exposure and melanocytic naevi in young Australian children. *Lancet* **344,** 1529–1532 (1994).
6. Dhomen, N. *et al.* Oncogenic Braf induces melanocyte senescence and melanoma in mice. *Cancer Cell* **15,** 294–303 (2009).
7. Pontén, F., Berne, B., Ren, Z. P., Nister, M. & Ponten, J. Ultraviolet light induces expression of p53 and p21 in human skin: effect of sunscreen and constitutive p21 expression in skin appendages. *J. Invest. Dermatol.* **105,** 402–406 (1995).
8. Rudolph, P., Tronnier, M., Menzel, R., Moller, M. & Parwaresch, R. Enhanced expression of Ki-67, topoisomerase IIα, PCNA, p53 and p21$^{WAF1/Cip1}$ reflecting proliferation and repair activity in UV-irradiated melanocytic nevi. *Hum. Pathol.* **29,** 1480–1487 (1998).
9. Pollock, P. M. *et al.* High frequency of *BRAF* mutations in nevi. *Nature Genet.* **33,** 19–20 (2002).
10. Thomas, N. E. *et al.* Number of nevi and early-life ambient UV exposure are associated with *BRAF*-mutant melanoma. *Cancer Epidemiol. Biomarkers Prev.* **16,** 991–997 (2007).
11. Vultur, A. & Herlyn, M. SnapShot: melanoma. *Cancer Cell* **23,** 706 (2013).
12. Pfeifer, G. P., You, Y. H. & Besaratinia, A. Mutations induced by ultraviolet light. *Mutat. Res.* **571,** 19–31 (2005).
13. Kato, S. *et al.* Understanding the function–structure and function–mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc. Natl Acad. Sci. USA* **100,** 8424–8429 (2003).
14. Song, H., Hollstein, M. & Xu, Y. p53 gain-of-function cancer mutants induce genetic instability by inactivating ATM. *Nature Cell Biol.* **9,** 573–580 (2007).
15. Willis, A., Jung, E. J., Wakefield, T. & Chen, X. Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene* **23,** 2330–2338 (2004).
16. Olive, K. P. *et al.* Mutant *p53* gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119,** 847–860 (2004).
17. Dankort, D. *et al.* Braf^{V600E} cooperates with *Pten* loss to induce metastatic melanoma. *Nature Genet.* **41,** 544–552 (2009).
18. Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150,** 251–263 (2012).
19. Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic *RAC1* mutations in melanoma. *Nature Genet.* **44,** 1006–1014 (2012).
20. Patton, E. E. & Zon, L. I. Taking human cancer genes to the fish: a transgenic model of melanoma in zebrafish. *Zebrafish* **1,** 363–368 (2005).
21. Goel, V. K. *et al.* Melanocytic nevus-like hyperplasia and melanoma in transgenic BRAFV600E mice. *Oncogene* **28,** 2289–2298 (2009).
22. Bald, T. *et al.* Ultraviolet-radiation-induced inflammation promotes angiotropism and metastasis in melanoma. *Nature* **507,** 109–113 (2014).
23. Noonan, F. P. *et al.* Neonatal sunburn and melanoma in mice. *Nature* **413,** 271–272 (2001).
24. Luo, C. *et al.* Loss of ARF sensitizes transgenic BRAF^{V600E} mice to UV-induced melanoma via suppression of XPC. *Cancer Res.* **73,** 4337–4348 (2013).
25. Zaidi, M. R. *et al.* Interferon-γ links ultraviolet radiation to melanomagenesis in mice. *Nature* **469,** 548–553 (2011).
26. Green, A. *et al.* Daily sunscreen application and betacarotene supplementation in prevention of basal-cell and squamous-cell carcinomas of the skin: a randomised controlled trial. *Lancet* **354,** 723–729 (1999).
27. Green, A. C., Williams, G. M., Logan, V. & Strutton, G. M. Reduced melanoma after regular sunscreen use: randomized trial follow-up. *J. Clin. Oncol.* **29,** 257–263 (2011).
28. Goldenhersh, M. A. & Koslowsky, M. Increased melanoma after regular sunscreen use? *J. Clin. Oncol.* **29,** e557–e558 (2011).
29. Planta, M. B. Sunscreen and melanoma: is our prevention message correct? *J. Am. Board Fam. Med.* **24,** 735–739 (2011).
30. Viros, A. *et al.* Improving melanoma classification by integrating genetic and morphologic features. *PLoS Med.* **5,** e120 (2008).

Author Contributions A.V., B.S.-L. and R.M. designed the study, analysed the data and wrote the paper. M.P. designed and performed experiments and analysed data. S.J.F. designed and performed bioinformatics analysis and analysed data. K.H., J.R., M.R.G., M.C. and N.D. performed experiments. S.E. validated WES SNVs.

Author Information Exome sequence and aCGH data have been deposited in the European Nucleotide Archive under study accession number PRJEB6330. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M. (richard.marais@cruk.manchester.ac.uk).

METHODS

Animal procedures. All procedures involving animals were approved by the Animal Ethics Committees of the Institute of Cancer Research and Cancer Research UK Manchester Institute in accordance with National Home Office regulations under the Animals (Scientific Procedures) Act 1986 and according to the guidelines of the Committee of the National Cancer Research Institute[31]. Tamoxifen (Sigma-Aldrich T5648) was freshly prepared in 100% ethanol. For genotyping, genomic DNA was prepared from tail biopsies and PCR was performed using the primers previously described[6]. Groups of >10 animals per cohort were based on historical precedent to provide statistically significant survival rates in Kaplan–Meier analysis. No randomization or blinding was required. Animals were killed when their tumour burden reached ethical limits, if they showed signs of ill health or distress, or after a maximum of 24 months on the study.

BRAF(V600E) expression and UVR treatments. BRAF(V600E) was expressed in female C57BL/6 mouse melanocytes by topical application of tamoxifen to the shaved backs of mice at ~2 months of age. A month later mice were anaesthetized and the whole backs were shaved. Half of the shaved backs were covered vertically with a UVR-proof cloth, to provide an internal control for non-UVR-exposed skin. Each animal was treated with $160\,mJ\,cm^{-2}$ UVA/UVB or $15\,J\,cm^{-2}$ UVA (Waldmann UV181 lamps fitted with UV 6 or UV A-1 tubes) each week for up to 6 months. For the sunscreen protection experiments, mice were treated with Sunsense Milk Sunscreen SPF 50 as recommended by the US Food and Drug Administration ($2.2\,mg\,cm^{-2}$) by topical application 30 min before UVR exposure.

Histology and immunohistochemistry. For immunohistochemistry, samples were processed as described[6]. Briefly, sections were deparaffinized with xylene and hydrated with a series of graded alcohol washes. Sections were microwaved in citrate buffer (pH 6) for antigen retrieval and rinsed in PBS washes. Sections were blocked in 1% BSA in PBS, incubated with 1:200 dilutions of anti-S100 (Dako), 1:100 dilutions of anti-Ki-67 (Santa Cruz), 1:200 dilutions of mouse anti-HMB-45/MelanA (Abcam), 1:100 Tp53 (Abcam) antibodies. Appropriate negative and positive controls for specificity of staining were included and images were captured using a Leica DM4000 B LED automated upright microscope system.

Naevus quantification. For each animal, at least three samples of skin were collected. The skin sections were cut perpendicular to the longitudinal axis of the animals to span the exposed and non-exposed areas of the tamoxifen-treated back. Two independent observers (observers 1 and 2, A.V. and M.C. respectively) performed naevus quantification blinded for animal group. For each animal, naevi were quantified in three different skin sections of 1.8 mm width each of UVR-exposed and protected skin. Significance was tested with the WRST. The size of the naevus was measured by a single observer recording the largest diameter for each naevus in 10.8 mm skin of UVR-exposed and non-exposed protected skin and significance was tested using the Mann–Whitney U test. Four animals were scored.

Trp53 quantification. We quantified Trp53 by scoring nuclear staining in 500 consecutive keratinocytes for each mouse in UVR-exposed, protected, and UVR plus sunscreen-treated skin. Four animals were scored.

Sunburn cell quantification. We quantified sunburn cells in 1 mm of interfollicular basal keratinocytes for each mouse in UVR-exposed, protected, and UVR plus sunscreen-treated skin. Four animals were scored.

Immunofluorescence. Three-millimetre sections of formalin-fixed paraffin-embedded (FFPE) material were used. Slides were dewaxed and antigen retrieval was performed using citrate buffer pH 6 followed by blocking in PBS-Tween 0.1% plus 1% BSA for 15 min and overnight incubation with anti-Ki-67 (Santa Cruz sc-7846) and S100 antibody (1:100 in PBS plus 1% BSA). Antibody detection was performed using AlexaFluor-conjugated secondary antibodies (Invitrogen). Slides were counterstained with 4′,6-diamidino-2-phenylindole (DAPI). Samples were analysed and pictures of the mid-dermis were taken using a Leica confocal scanning microscope (Leica Microsystems).

Tumour analysis. Mouse genomic DNA was prepared for aCGH and WES. One-hundred and seven randomly selected SNVs were validated by Sanger sequencing to reveal a true positive rate of 93.5%.

WES and aCGH. Seventy-five manually dissected sections of 20 μm snap-frozen tissue samples with an estimated tumour cell percentage of at least 80% were used for DNA extraction. Normal DNA was isolated from liver or kidney. DNA extraction was performed using a DNeasy blood and tissue kit (Qiagen). Quality assessment was performed by NanoDrop and Qubit. For aCGH, 2 μg of normal and tumour DNA were hybridized to Nimblegen HX3 MM9 whole-genome CGH arrays (100718_MM9_WG_CGH). Normalization, window averaging and segmentation were conducted using the standard Nimblegen aCGH analysis pipeline. The BioConductor package aCGH was used to summarize and plot copy number alterations. For WES, 15 tumours from UVR-exposed areas, 15 tumours from animals not exposed to UVR and 11 tumours from animals exposed to UVR and treated with sunscreen were included. Matched normal DNA samples from liver or kidney were obtained. DNA was extracted from frozen materials. Exome capture was performed using the Agilent SureSelect Mouse exome capture kit with 3 μg of genomic DNA, according to the manufacturer's instructions. Sequencing was performed on Illumina GAIIx and HiSeq machines to produce 74 bp paired-end reads and quality control was conducted using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the NCBI m37 reference genome using BWA[32] (version 0.5.9-r16; default parameters). Duplicate reads were marked by Picard (http://picard.sourceforge.net/index.shtml) and base quality score recalibration and local realignment around indels were performed using GATK[33] (version 1.2-24-g6478681). Pileup files of on-target high mapping quality bases (read mapping quality ≥30) were generated using Samtools[34] (version 0.1.18) and somatic SNVs were called using Varscan[35] (version 2.2.8) methods somatic (parameters: -min-var.-freq 0.08, -p-value 0.05, -min-avg-qual 20) and processSomatic at positions with coverage ≥10 in the normal tissue and tumour. Likely false positives were removed using predetermined filters[35]. The somatic mutational status of *Trp53* in all tumour samples was also verified using Somaticsniper[36]. No mutations in *Trp53* in the non-UVR melanoma samples were identified by either method. Somatic variants were annotated by the Ensembl Variant Effect Predictor[37] using Ensembl version 67 and variants present in dbSNP were excluded.

For Sanger sequencing validation, genomic DNA was amplified by PCR to confirm selected mutations. The products were directly sequenced using dye-terminator chemistry as previously described[38]. Sequences were visualized using Sequencher software. Primer sequences are available upon request. One-hundred and seven randomly selected SNVs were validated by Sanger sequencing, to reveal a true positive rate of 93.5%.

31. Workman, P. *et al.* Guidelines for the welfare and use of animals in cancer research. *Br. J. Cancer* **102**, 1555–1577 (2010).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
34. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
35. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
36. Larson, D. E. *et al.* SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* **28**, 311–317 (2012).
37. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
38. Turajlic, S. *et al.* Whole genome sequencing of matched primary and metastatic acral melanomas. *Genome Res.* **22**, 196–207 (2012).
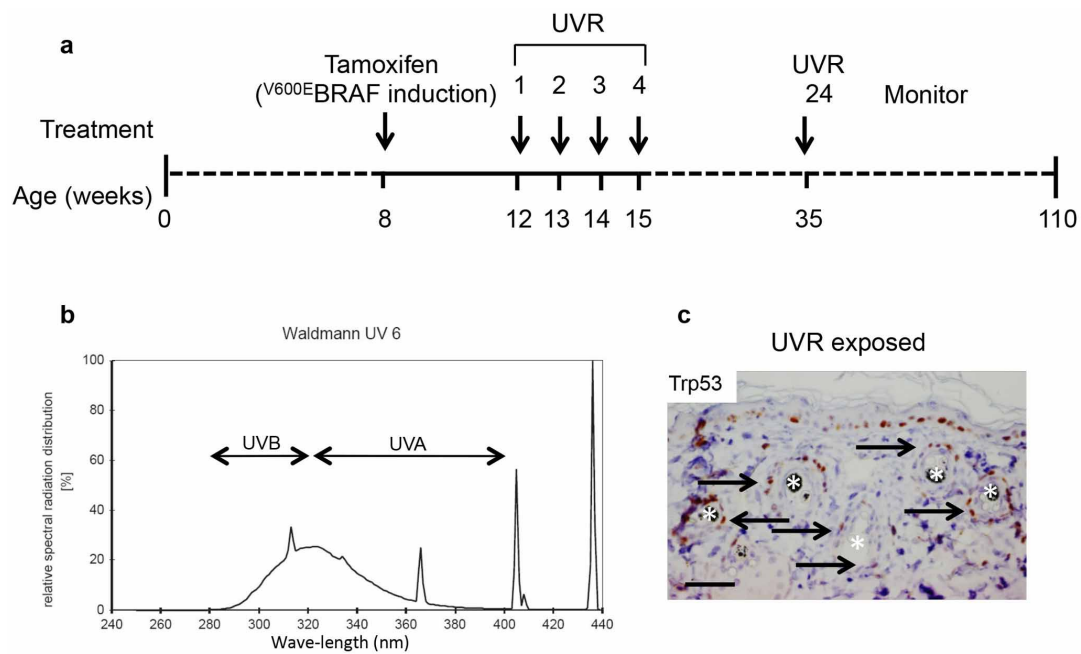
Extended Data Figure 1 | Methodology and UVR spectrum. a, Schematic representation of experimental schedule, showing induction of BRAF(V600E) by tamoxifen at ~8 weeks of age followed by weekly exposure to UVR starting 4 weeks later, and for up to 6 months. **b**, Graph showing spectral radiation distribution for the Waldmann UV6 lamp used in these studies. The UVA and UVB regions are indicated. **c**, Photomicrograph of Trp53 staining in interfollicular and follicular keratinocytes (arrows) 24 h after UVR exposure. Asterisks indicate hair follicles. Scale bar, 50 μm.
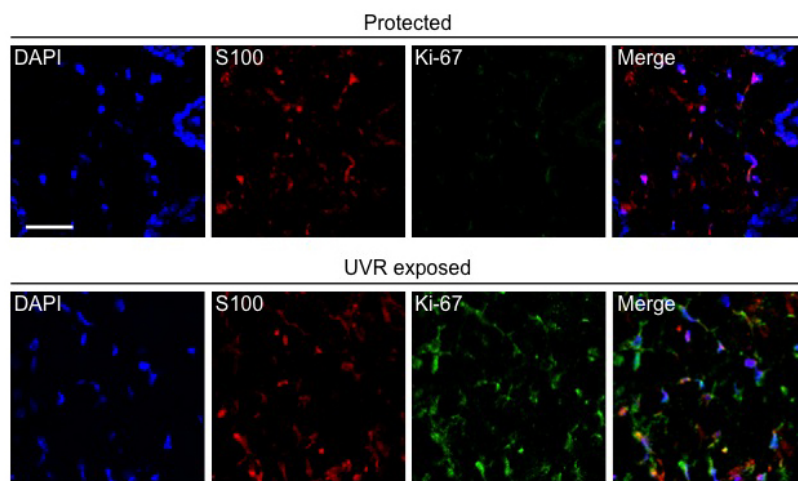
Extended Data Figure 2 | UVR induced melanocytic proliferation in BRAF(V600E) mice. Photomicrographs showing 4′,6-diamidino-2-phenylindole (DAPI), S100 and Ki-67 immunofluorescence staining, together with a merged image in cloth-protected and UVR-exposed areas of the mid-dermis from BRAF(V600E) mice 72 h after UVR exposure. The experiment was repeated in 3 different skin sections of 5 animals. Scale bar, 30 µm.
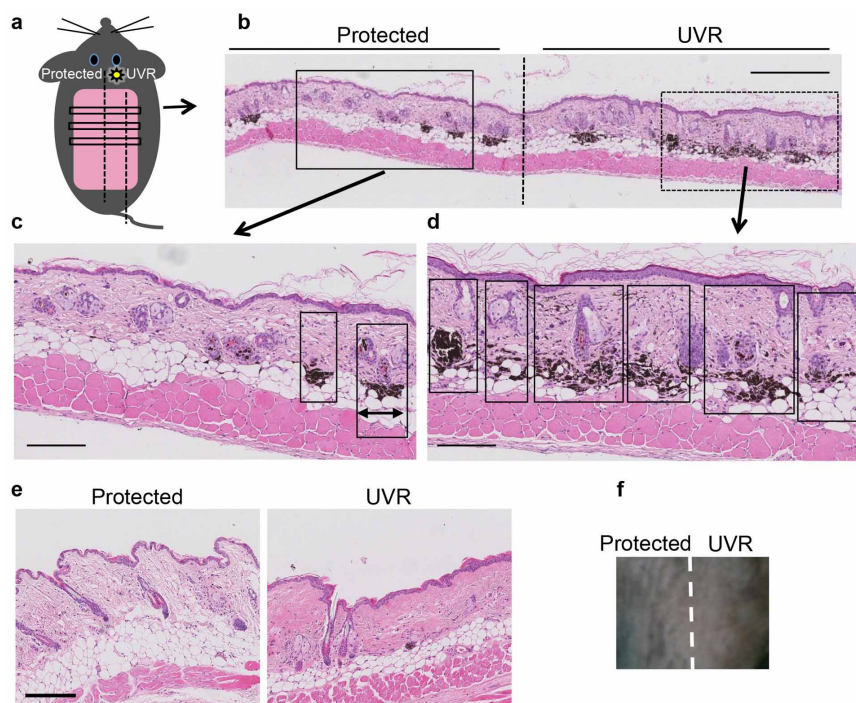
Extended Data Figure 3 | UVR induces naevogenesis in BRAF(V600E) mice. a, Schematic representation of experimental processing of mouse skin. Three skin sections (long rectangular boxes) were cut perpendicular to the longitudinal axis of the mice to span the UVR-exposed and protected areas (as marked by the dotted line) of the tamoxifen-treated shaved area (pink). **b,** Photomicrograph of H&E-stained representative skin from the protected and UVR-exposed skin of a tamoxifen-treated BRAF(V600E) mouse subjected to UVR exposure and examined at 7 days ($n = 5$). Dotted line: UVR treatment line demarcation. Scale bar, 0.4 mm. **c,** Photomicrograph of H&E-stained skin from the boxed black area in **b,** showing the cloth-protected skin of a tamoxifen-treated BRAF(V600E) mouse subjected to UVR exposure and examined at 7 days. Black boxes show individual naevi. Double-headed arrow shows example of maximum diameter of a single naevus. Scale bar, 300 μm. **d,** Photomicrograph of H&E-stained skin from the dashed boxed black area in **b,** showing the UVR-exposed skin of a tamoxifen-treated BRAF(V600E) mouse subjected to UVR exposure and examined at 7 days. Black boxes show individual naevi. Scale bar, 300 μm. **e,** Photomicrograph of H&E-stained skin from the cloth-protected and UVR-exposed skin of a tamoxifen-treated CreERT2 control mouse subjected to weekly UVR exposure for 6 months. Scale bar, 300 μm. **f,** Photograph showing macroscopic appearance of protected and UVR-exposed skin from the tamoxifen-treated CreERT2 control mouse shown in **a.** Original magnification, ×2.
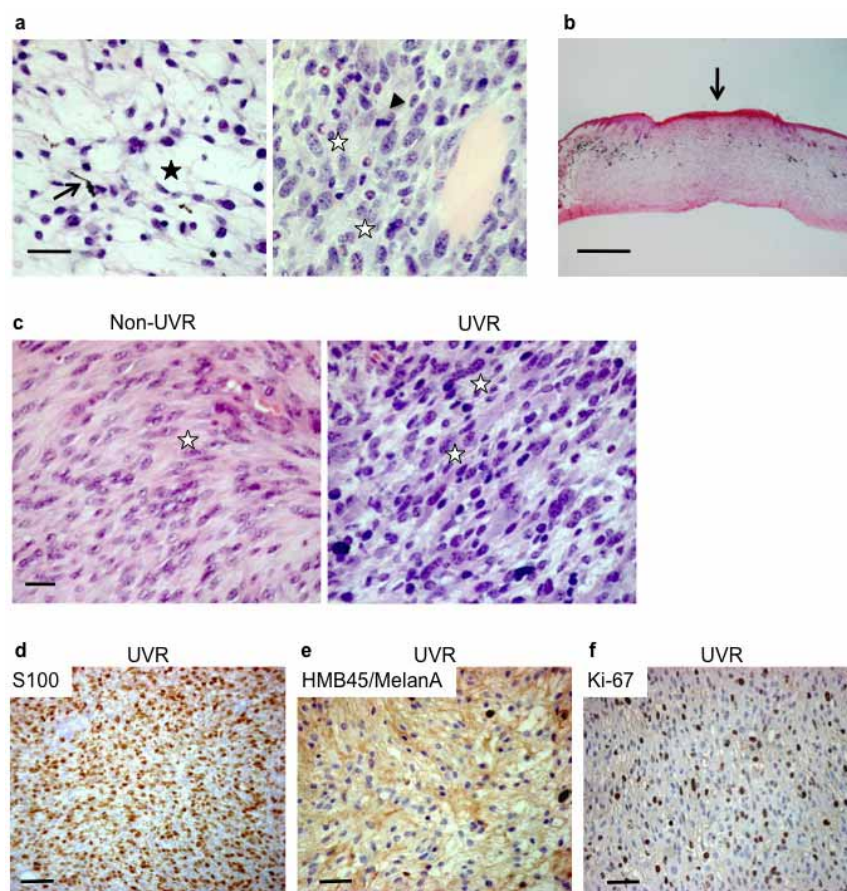
Extended Data Figure 4 | UVR-accelerated BRAF(V600E)-driven tumours have similar histology to BRAF(V600E)-driven tumours from non-UVR-exposed animals. a, Photomicrograph of an H&E-stained tumour from a UVR-treated BRAF(V600E) mouse highlighting the presence of atypical heterogeneous spindle dendritic and pigment-producing cells (black asterisks and arrow respectively, left panel) and atypical plump spindle cells and mitotic figures (white asterisks and black arrowhead respectively, right panel). Scale bar, 25 μm. **b,** Photomicrograph of an H&E-stained tumour from a UVR-treated BRAF(V600E) mouse highlighting a region of dermal ulceration (black arrow). Scale bar, 0.5 mm. **c,** Photomicrograph of H&E-stained tumours from non-UVR exposed (left panel) and UVR-exposed (right panel) BRAF(V600E) mice, showing the presence of nuclear pleomorphism (asterisks) in both tumours. Scale bar, 10 μm. **d,** Photomicrograph of an S100-stained tumour from a UVR-exposed BRAF(V600E) mouse. Scale bar, 100 μm. **e,** Photomicrograph of an HMB-45/MelanA-stained tumour from a UVR-exposed BRAF(V600E) mouse. Scale bar, 50 μm. **f,** Photomicrograph of a Ki-67-stained tumour from a UVR-exposed BRAF(V600E) mouse. Scale bar, 50 μm.
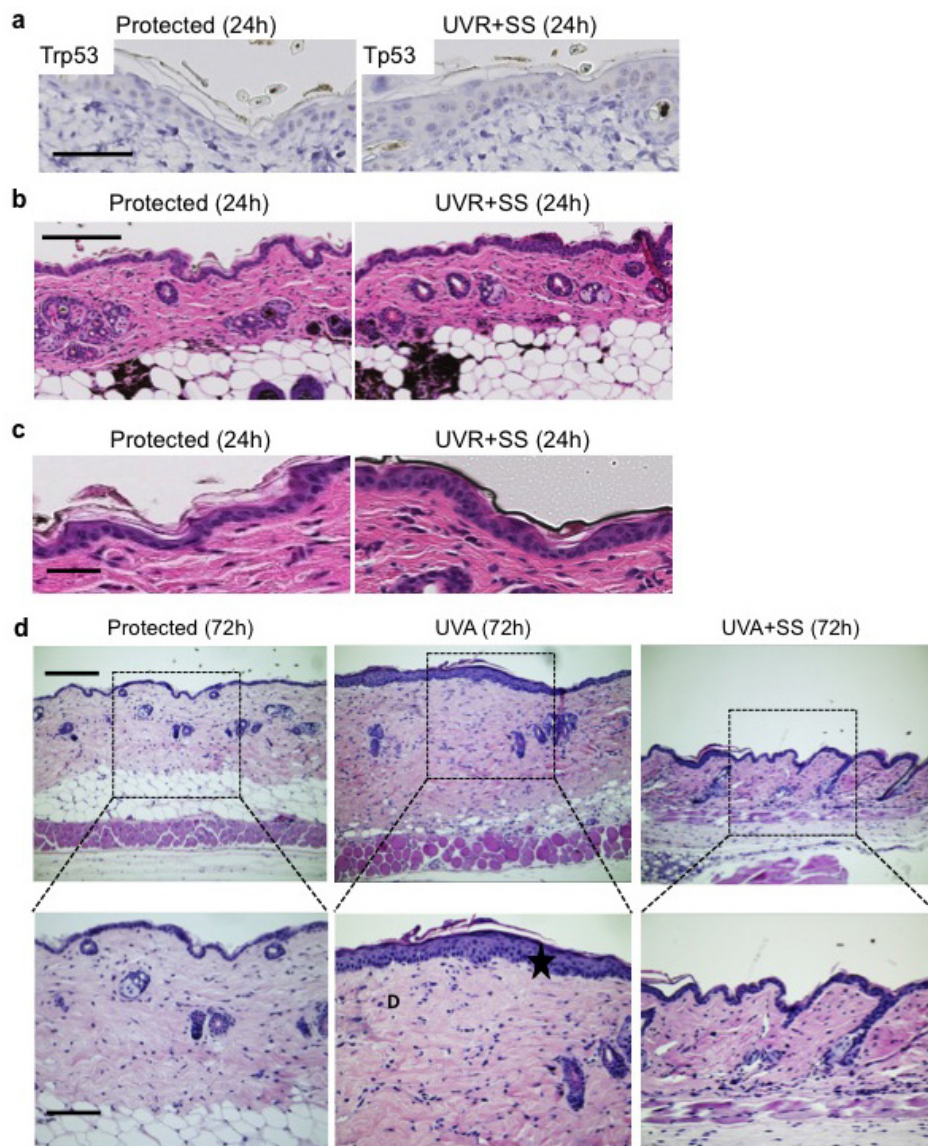
Extended Data Figure 5 | Sunscreen blocks the short-term effects of UVR exposure. a, Photomicrograph showing lack of Trp53 staining in the cloth-protected (Protected) and UVR-exposed, sunscreen-protected (UVR+SS) skin of a BRAF(V600E) mouse after 24 h of UVR exposure ($n = 5$). Scale bar, 40 μm. **b**, Photomicrograph of H&E-stained skin from the cloth protected (Protected) and UVR-exposed, sunscreen-protected (UVR+SS) regions of a BRAF(V600E) mouse after 24 h of UVR exposure. Scale bar, 300 μm. **c**, High-magnification photomicrograph of H&E-stained skin from the cloth protected (Protected) and UVR-exposed sunscreen-protected (UVR+SS) regions of a BRAF(V600E) mouse after 24 h of UVR exposure showing absence of apoptotic keratinocytes. **d**, Photomicrograph of H&E-stained skin from a cloth-protected (Protected; left), UVA-exposed (UVA; centre), and UVA-exposed, sunscreen-protected (UVA+SS; right) skin of a mouse after 72 h of UVR exposure. Scale bar, 300 μm ($n = 5$). The photomicrographs below the main images show areas of higher magnification of the images above. Asterisk indicates epidermal hypertrophy; D indicates plump collagen bundles densely located in the dermis. Scale bar, 150 μm.
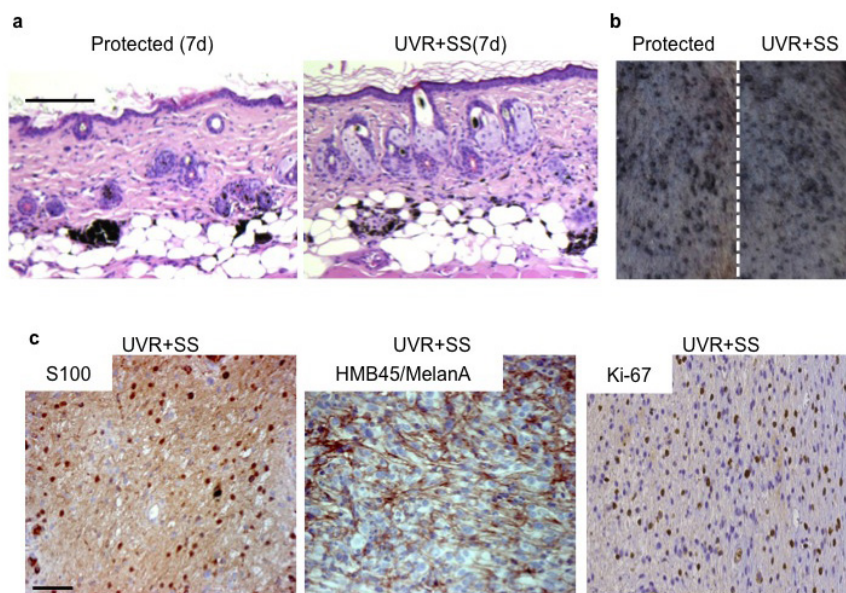
Extended Data Figure 6 | UVA-induced epidermal and dermal thickening after UVA radiation is abrogated by the use of sunscreen.
a, Photomicrographs of H&E-stained epidermis from the cloth-protected (Protected) and UVR-exposed, sunscreen-protected (UVR+SS) skin of a BRAF(V600E) mouse 7 days (d) after UVR exposure showing the similarity in the number and size of naevi in the two regions ($n = 5$). Scale bar, 300 μm.

b, Photographs showing the macroscopic appearance of the cloth-protected (Protected) and UVR-exposed, sunscreen-protected (UVR+SS) skin from a BRAF(V600E) mouse 5 months after UVR. **c**, Photomicrographs of S100, HMB-45/MelanA- and Ki-67-stained tumours from UVR-exposed, sunscreen-protected (UVR+SS) mice. Scale bar, 50 μm.
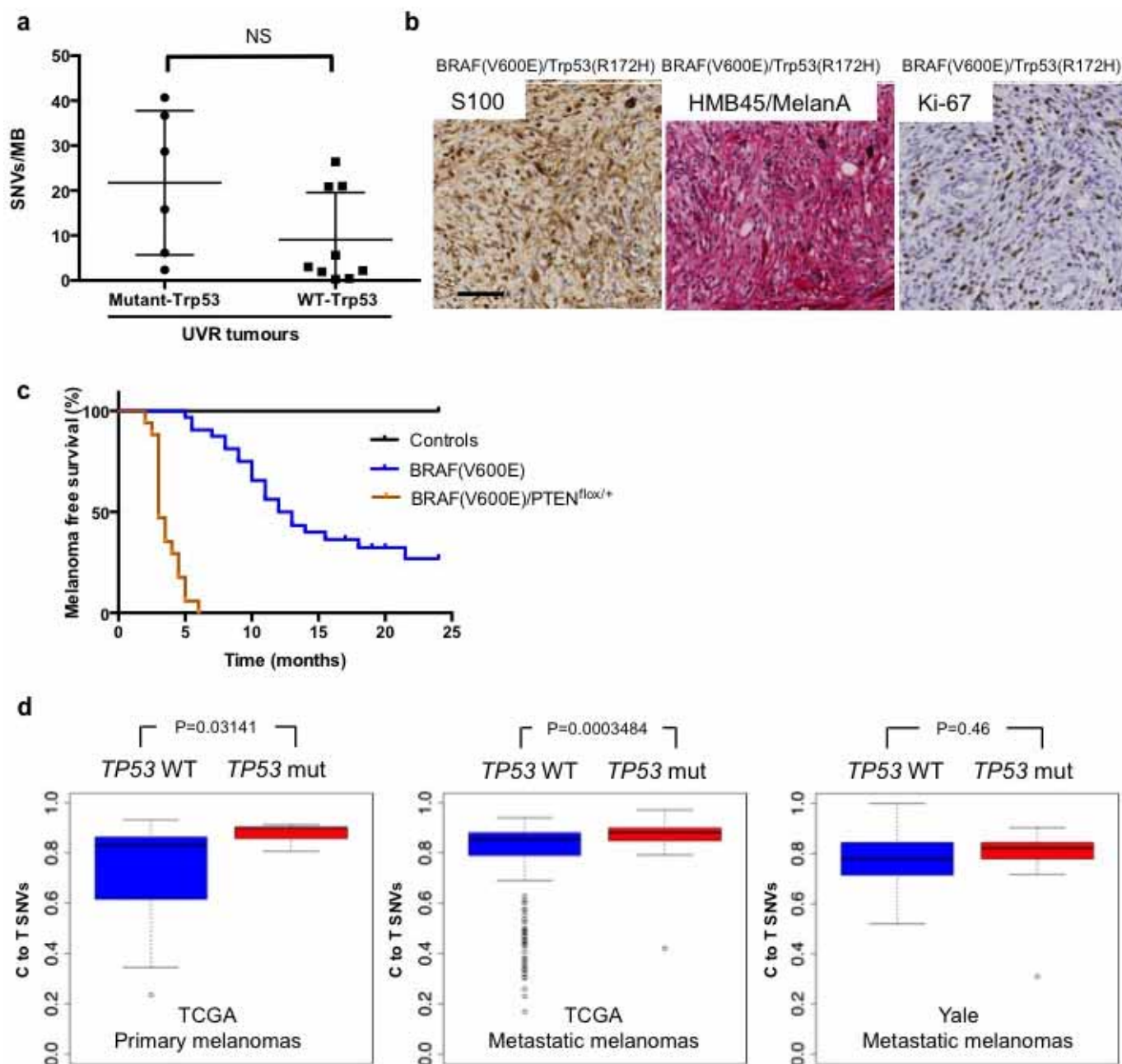
Extended Data Figure 7 | Tumours with a Trp53/TP53 mutation. a, Boxplot graph showing median number of SNVs per Mb (SNVs per Mb) in UVR-exposed mutant *Trp53*, or UVR-exposed wild-type *Trp53* (WT-Trp53) tumours. NS, not significant (Mann–Whitney test). **b,** Photomicrographs of S100, HMB-45/MelanA- and Ki-67-stained tumour sections from a BRAF(V600E)/Trp53(R172H) mouse. Scale bar, 50 μm. **c,** Kaplan–Meier plot showing melanoma-free survival in control mice (CreERT2; black line);

BRAF(V600E) mice (BRAF(V600E) blue line); and BRAF(V600E) mice crossed with PTEN $^{flox/+}$ mice (BRAF(V600E)/ PTEN$^{flox/+}$; orange line). **d,** Boxplot graphs showing proportion of C-to-T (G-to-A) transitions in human primary melanomas from the TCGA data set (left panel), human metastatic melanomas from the TCGA data set (middle) and human metastatic melanomas from the Yale University data set (right) harbouring wild-type *TP53* (*TP53* WT) or mutations in *TP53* (*TP53* mut); WSRT.

Extended Data Table 1 | UVR induces Trp53 and apoptosis in epidermal keratinocytes

	Protected	UVR exposed	UVR exposed + sunscreen
Trp53 positive	0	41(35-60)%	0
Apoptotic cells	0	17(9-27)/mm	0

The table shows quantification of Trp53 and apoptotic keratinocytes (sunburn cells) in keratinocytes from cloth-protected, UVR-exposed and UVR-exposed sunscreen-protected skin 24 h after exposure. The data are presented as percentage of positive cells, with range in brackets, and show cells presenting Trp53-positive nuclei (Trp53 positive) and average number of apoptotic keratinocytes (Apoptotic cells) in the interfollicular epidermis for each condition.

Extended Data Table 2 | Quantification of naevi

	Protected	UVR	p value
Naevi Number Obs 1	7(2-10)	11(6-14)	0.003*
Naevi Number Obs 2	5(3-7)	9(6-11)	0.0001*
Naevi Size Obs 1	0.9(0.2-1.5)	1.3(0.4-2.3)	0.0002**

Average number, with range in brackets, of naevi present in 1.8 mm protected and 1.8 mm UVR-exposed skin of animals 7 days after UVR treatment. Naevi number quantification was performed by two independent observers (Obs1, Obs2). Average size of naevi (range, arbitrary units) observed in 10.8 mm protected and 10.8 mm UVR-exposed skin of animals 7 days after UVR treatment.
* WRST.
** Mann–Whitney U test.

Extended Data Table 3 | Summary of *Trp53* mutations in UVR-exposed BRAF(V600E) mouse tumours and the *TP53* mutation context in humans

a

Tumour	Mouse	Sanger validation	UV induced mutation	Mutation context	IHC	Human	Human somatic mutation frequency	Mutated in human melanoma
UVR 1	R245C	Yes	C/T	ACCGC	Nuclear 3+	R248	1887	Yes
UVR 2	H39Y	Yes	C/T	CTCAC	Nuclear 3+	Q38	4	Not found
UVR 5	R270C	Yes	C/T	TTCGT	Nuclear 3+	R273	1790	Yes
UVR 6	C272G	Yes	T/G	TTTGT	Nuclear 2+	C275	175	Yes
UVR 12	S124F	Yes	C/T	CTCTC	Nuclear 3+	S127	64	Yes
UVR 15	R245C	Yes	C/T	ACCGC	Nuclear 3+	R248	1887	Yes
UVR SS 8	R270C	Yes	C/T	TTCGT	NA	R273	1790	Yes
BRAF(V600E)/ TP53(R172H	R172H	NA	NA	NA	Nuclear 2+	R175H	1303	Yes

b

Human mutation	SIFT	Ability to transactivate (% of WT activity)		
		WAF1/p21	MDM2	BAX1
S127F	Deleterious	20.6	20.1	14.3
R248C	Deleterious	0	0.6	0
R273C	Deleterious	1.5	0	2.7
C275G	Deleterious	1	0	0
R175H	Deleterious	20.5	17.6	10.5

c

	TCGA n (%)	Broad n (%)	Yale n (%)	Total n (%)
Total *TP53* mutations	51 (100)	27 (100)	9 (100)	87 (100)
TP53 mutations with a UVR signature	34 (66.6667)	13 (48.1482)	5 (55.5556)	52 (59.7701)

a, The table summarizes the *Trp53* mutations in UVR-exposed mouse tumours, shows Sanger sequencing validation, the UVR-induced mutations, mutation context and records the intensity of Trp53 staining by immunohistochemistry. The table also shows the corresponding human codon and the frequency of somatic mutations at those codons in human cancer (http://p53.iarc.fr/TP53GeneVariations.aspx) and if this codon is mutated in human melanoma. NA, not available. **b**, The table shows the *Trp53* mutations in UVR-exposed BRAF(V600E) tumours, the corresponding human codon, the functional classification based on the SIFT program (SIFT) and the promoter-specific transcriptional activity measured in yeast functional assays and expressed as percentage of wild-type (WT) activity[13]. **c**, *TP53* mutations occurring at the 3′ end of pyrimidine dimers in human TCGA, Broad Institute and Yale University databases. *n*, number of human samples with a *TP53* mutation (top row) and *TP53* mutation at the 3′ end of pyrimidine dimers (bottom row). % indicates column percentage.

LETTER

Activation and repression by oncogenic MYC shape tumour–specific gene expression profiles

Susanne Walz[1]*, Francesca Lorenzin[1]*, Jennifer Morton[2], Katrin E. Wiese[1], Björn von Eyss[1], Steffi Herold[1], Lukas Rycak[3], Hélène Dumay-Odelot[4], Saadia Karim[2], Marek Bartkuhn[5], Frederik Roels[6], Torsten Wüstefeld[7], Matthias Fischer[6], Martin Teichmann[4], Lars Zender[7,8], Chia-Lin Wei[9], Owen Sansom[2], Elmar Wolf[1,10]* & Martin Eilers[1,11]*

In mammalian cells, the MYC oncoprotein binds to thousands of promoters[1–4]. During mitogenic stimulation of primary lymphocytes, MYC promotes an increase in the expression of virtually all genes[1]. In contrast, MYC-driven tumour cells differ from normal cells in the expression of specific sets of up- and downregulated genes that have considerable prognostic value[5–7]. To understand this discrepancy, we studied the consequences of inducible expression and depletion of MYC in human cells and murine tumour models. Changes in MYC levels activate and repress specific sets of direct target genes that are characteristic of MYC-transformed tumour cells. Three factors account for this specificity. First, the magnitude of response parallels the change in occupancy by MYC at each promoter. Functionally distinct classes of target genes differ in the E-box sequence bound by MYC, suggesting that different cellular responses to physiological and oncogenic MYC levels are controlled by promoter affinity. Second, MYC both positively and negatively affects transcription initiation independent of its effect on transcriptional elongation[8]. Third, complex formation with MIZ1 (also known as ZBTB17)[9] mediates repression of multiple target genes by MYC and the ratio of MYC and MIZ1 bound to each promoter correlates with the direction of response.
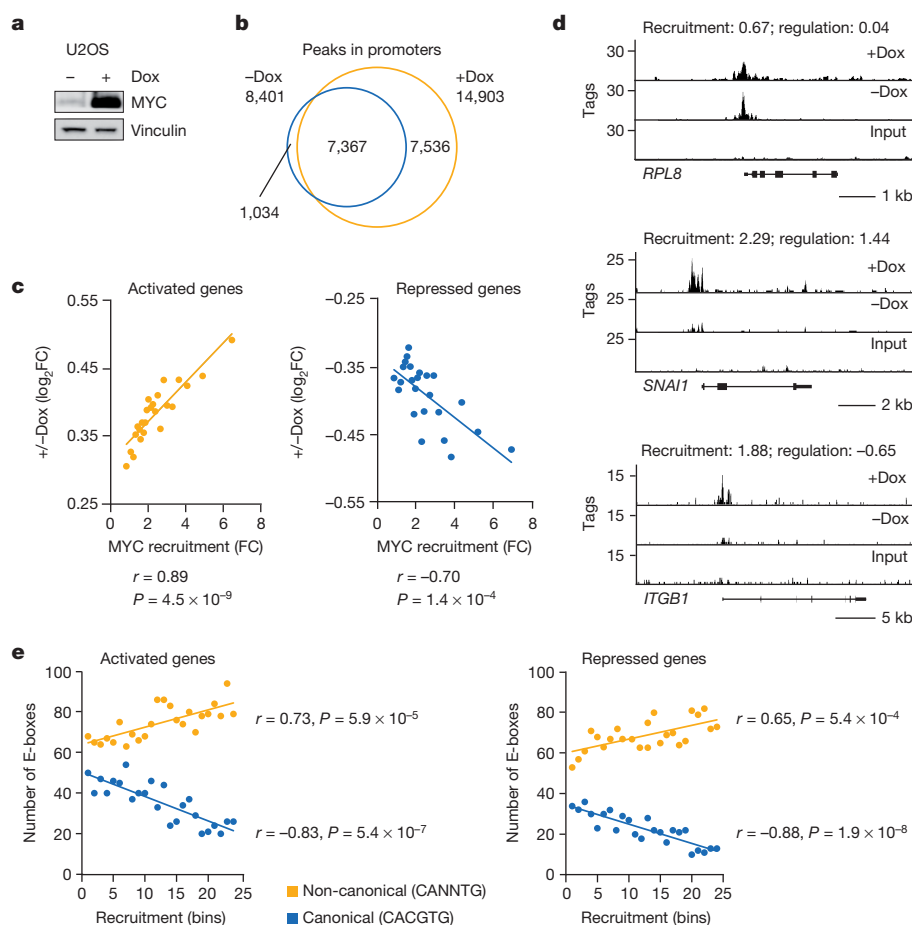
Figure 1 | Oncogenic levels of MYC establish tumour-cell-specific gene expression patterns. a, MYC levels in U2OS cells before and after induction with doxycycline (Dox). Vinculin was used as a loading control. **b**, Number of MYC-binding sites in Pol II promoters. **c**, Diagram depicting change in gene expression versus change in occupancy upon induction of MYC for 9,550 promoters. Activated and repressed genes were divided into 25 bins each. P values, two-tailed Student's t-test. **d**, Examples of MYC recruitment to different promoters. **e**, Number of E-boxes in MYC-binding sites stratified according to change in MYC occupancy ($\log_2$FC). P values, two-tailed Student's t-test.

[1]Theodor Boveri Institute, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany. [2]CRUK Beatson Institute, Garscube Estate, Switchback Road, Glasgow G61 1BD, UK. [3]Institute for Molecular Biology and Tumor Research (IMT), Emil-Mannkopff-Str.2, 35033 Marburg, Germany. [4]University of Bordeaux, IECB, ARNA laboratory, Equipe Labellisée Contre le Cancer, 33600 Pessac, France. [5]Institute for Genetics, Justus-Liebig-University, Heinrich-Buff-Ring 58, 35390 Giessen, Germany. [6]University Children's Hospital of Cologne, and Cologne Center for Molecular Medicine (CMMC), University of Cologne, Kerpener Str. 62, 50924 Cologne, Germany. [7]University Hospital Tübingen, Division of Translational Gastrointestinal Oncology, Department of Internal Medicine I, Otfried-Mueller-Strasse 10, 72076 Tübingen, Germany. [8]Translational Gastrointestinal Oncology Group within the German Center for Translational Cancer Research (DKTK), German Cancer Research Center (DKFZ), 69121 Heidelberg, Germany. [9]DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA. [10]Rudolf Virchow Center/DFG Research Center for Experimental Biomedicine, University of Würzburg, Josef-Schneider-Str.2, 97080 Würzburg, Germany. [11]Comprehensive Cancer Center Mainfranken, University of Würzburg, Josef-Schneider-Str. 6, 97080 Würzburg, Germany.
*These authors contributed equally to this work.

We expressed doxycycline-inducible MYC in U2OS cells that express low levels of endogenous MYC. Immunoblot and quantitative polymerase chain reaction with reverse transcription (RT–qPCR) analyses demonstrated a large increase in *MYC* messenger RNA leading to an approximately tenfold increase in MYC protein upon addition of doxycycline (Fig. 1a and Extended Data Fig. 1a, b). Addition of doxycycline to growing cells led to marginal alterations in cell size and the percentage of 5-bromodeoxyuridine (BrdU)-positive cells (Extended Data Fig. 1c, d). Upon prolonged exposure to doxycycline cells underwent apoptosis (Extended Data Fig. 1e)[10]. Induction of MYC had no significant effect on RNA or mRNA content per cell (Extended Data Fig. 1f). Chromatin immunoprecipitation (ChIP)-sequencing identified 20,014 peaks for endogenous MYC (Extended Data Table 1 and Extended Data Fig. 2e). MYC binding was enriched at promoters transcribed by RNA polymerases II (Pol II) and III, defining 8,401 MYC-bound Pol II promoters (Fig. 1b). Upon addition of doxycycline, the number of MYC-bound loci increased to 45,645, of which 14,903 were localized in promoters (Fig. 1b and Extended Data Fig. 3a). Consistent with previous observations, 3,656 MYC-binding sites were localized at enhancers (Extended Data Fig. 3b, c)[2].

RNA-sequencing showed that 462 genes were up- and 896 downregulated by MYC, of which 220 upregulated and 256 downregulated genes were direct target genes (Extended Data Fig. 3d). Gene set enrichment analysis (GSEA)[11] showed that MYC-regulated genes were highly similar to previously identified target genes (Extended Data Fig. 3e) and stratified human tumours according to *MYC* amplification (Extended Data Fig. 3f). Consistently, a linear support vector machine algorithm based on the set of MYC-regulated genes correctly classified 37 of 38 neuroblastomas as

harbouring a single copy or amplified *MYCN* gene (accuracy = 0.97 and Matthews correlation coefficient (mcc) = 0.95).

Short hairpin RNA (shRNA)-mediated depletion of MYC in HeLa cells, which express high levels of MYC, had no discernible effect on cell size or RNA levels per cell (Extended Data Fig. 4a–c). Using a cut-off of 1.5-fold change, depletion of MYC led to downregulation of 649 and upregulation of 608 genes. ChIP-sequencing identified 30,487 MYC-binding sites, 7,225 of which are located in Pol II promoters (see later and Extended Data Fig. 6a). Combining both data sets showed that depletion of MYC downregulated 242 and upregulated 162 direct target genes. GSEA showed that both up- and downregulated genes were highly similar to previously identified MYC-regulated genes, correlated with signatures of MYC-regulated genes in human tumours[12] (Extended Data Fig. 4d) and allowed stratification of lung tumours according to *MYC* amplification (Extended Data Fig. 4e). A support vector machine algorithm correctly identified 37 of 38 neuroblastomas as harbouring an amplified or single copy of the *MYCN* gene (accuracy = 0.97; mcc = 0.95). We concluded that oncogenic MYC levels establish gene expression patterns that are characteristic of human tumour cells which express high levels of *MYC* or *MYCN*.

ChIP-sequencing showed that the overall occupancy by MYC as well as the change in occupancy in response to doxycycline varied widely among promoters. The change in expression did not correlate with overall occupancy but was proportional to the change in occupancy of the respective promoter in response to doxycycline (Fig. 1c). Changes in MYC occupancy at enhancers had no consistent effect on expression of the neighbouring genes (Extended data Fig. 5a, b). Gene ontology (GO)-term
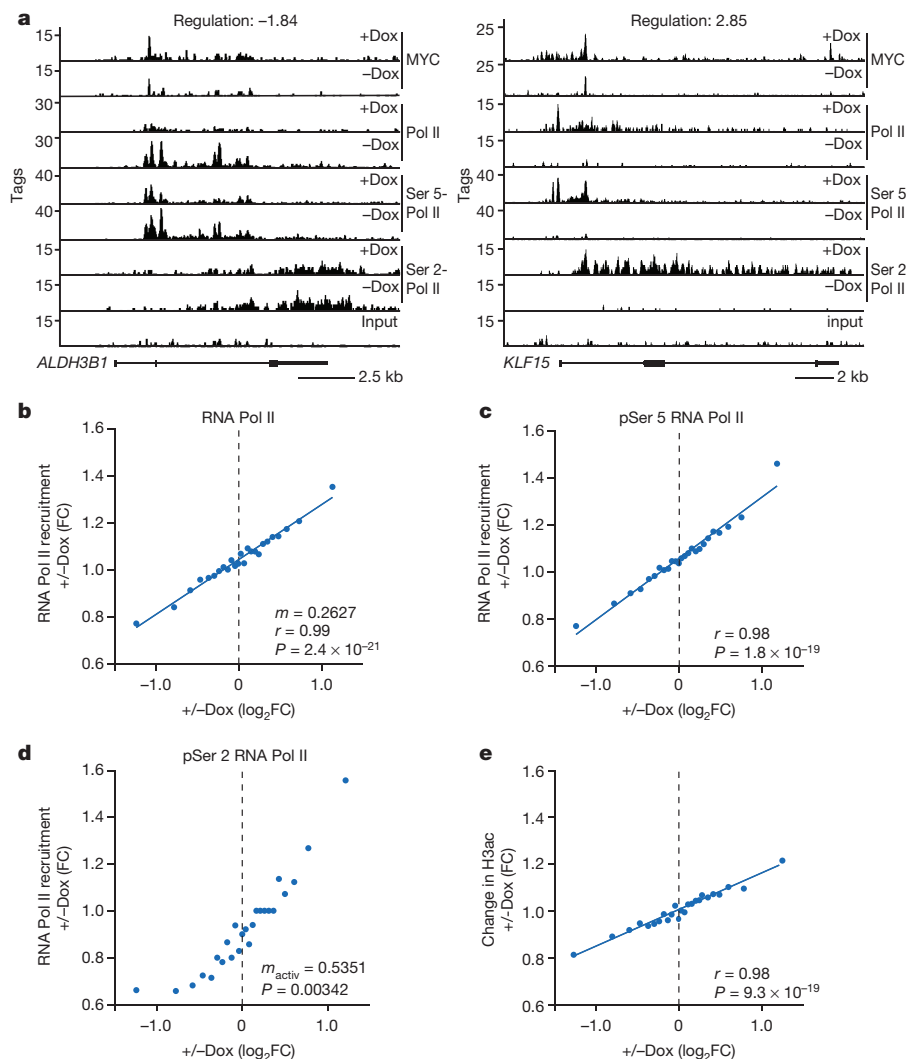
Figure 2 | MYC regulates recruitment of Pol II to target promoters. a, Examples of Pol II and MYC ChIP-sequencing traces. **b–d**, Changes in total Pol II (**b**), Ser 5-phosphorylated Pol II (**c**) and Ser 2-phosphorylated Pol II (**d**) occupancy at MYC-bound promoters in response to induction of MYC. Data are stratified by the MYC-induced change in expression. Each bin represents median values of 380 genes. *m*, slope of the linear regression; m_{activ}, slope of regression of activated genes; *r*, Pearson correlation coefficient. *P* values, two-tailed Student's *t*-test. **e**, Change in histone H3 acetylation (H3ac) at all MYC-bound promoters in response to MYC. Data are plotted as in **b–d**.

analysis showed that genes encoding proteins involved in ribosome biogenesis, translation and mitochondrial function were highly occupied in proliferating cells and neither occupancy nor expression increased further when MYC expression reached supra-physiological levels (Fig. 1d and Extended Data Fig. 5c). Target genes that mediate the oncogenic effects of MYC on cell migration, angiogenesis and metastasis showed strong changes in occupancy and expression (Fig. 1d and Extended Data Fig. 5c). The change in occupancy positively correlated with the fraction of non-consensus E-boxes (CANNTG) that are low-affinity MYC-binding sites (Fig. 1e)[12]. Promoters of genes that mediate functions in cell growth were enriched for consensus E-box sequences (CACGTG), which are high-affinity binding sites, and their fraction was negatively correlated with the change in MYC occupancy (Fig. 1e)[12]. We concluded that differences in binding affinity enable different levels of MYC to regulate functionally distinct classes of promoters (Extended Data Fig. 5d).

Transcriptional activation by MYC involves recruitment of TFIIH and P-TEFb (also known as CDK9) and an increase in transcriptional elongation[8,13–16]. We therefore performed ChIP-sequencing using antibodies

that recognize total Pol II, and Pol II phosphorylated at Ser 5 and Ser 2 of the carboxy-terminal domain (CTD), hallmarks of initiating and elongating Pol II, respectively (Fig. 2a)[17]. Changes in gene expression paralleled changes in the binding of total Pol II and of Ser 5-phosphorylated Pol II to the respective promoter (Fig. 2b, c). Changes in occupancy by Ser 2-phosphorylated Pol II at the transcription end site were higher than expected on activated and weaker than expected on repressed genes (Fig. 2d; $P = 0.00342$ for the difference in slope; analysis of covariance (ANCOVA)). A possible explanation is that regulation of Pol II recruitment by MYC is additive with its effect on elongation on activated genes but antagonizes it on repressed genes. MYC-induced changes in acetylation of histones H3 and H4 paralleled changes in gene expression (Fig. 2e and Extended Data Fig. 5e).

To understand how association with MIZ1 affects the response to MYC, we identified 6,086 MIZ1-binding sites in HeLa cells, of which 3,270 were located in Pol II promoters (Fig. 3a and Extended Data Figs 2a–d, 6a, b). Like MYC, binding of MIZ1 paralleled occupancy of Pol II and correlated with the presence of histone H3 trimethylated at lysine 4
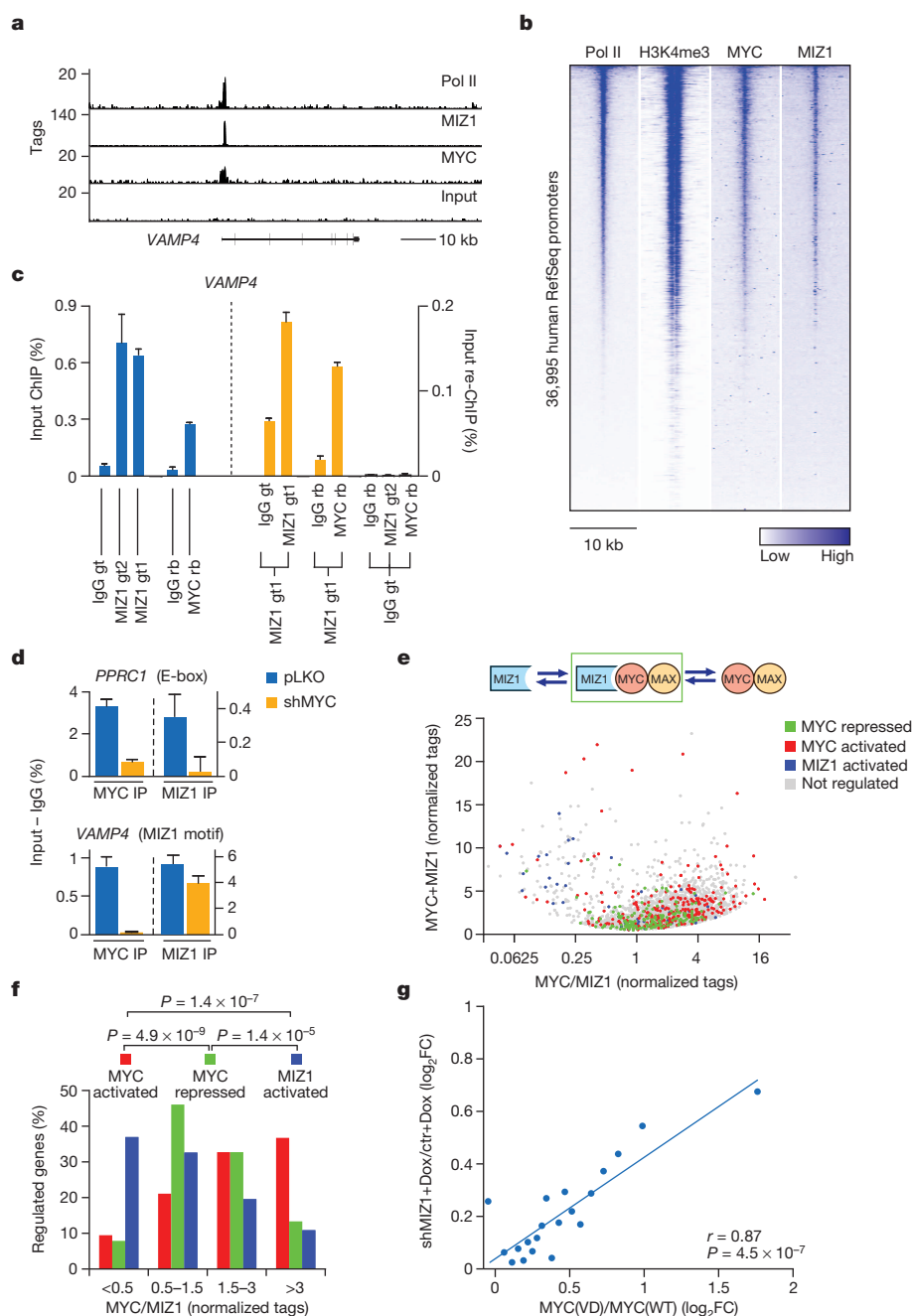
Figure 3 | Characterization of joint MYC/MIZ1-binding sites.
a, Example of ChIP-sequencing traces. **b**, Correlation of MYC and MIZ1 with Pol II binding and H3K4me3 at the transcription start site (TSS). **c**, Re-ChIP at the *VAMP4* promoter. Left, primary ChIP; right, re-ChIP. IgG gt, goat IgG; IgG rb, rabbit IgG. Error bars show standard deviation (s.d.) of technical replicates from a representative experiment ($n = 3$). **d**, Binding of MIZ1 and MYC in control and MYC-depleted cells to the *PPRC1* and *VAMP4* promoters. IP, immunoprecipitation. **e**, Sum of MYC- and MIZ1-binding tags versus ratio of MYC/MIZ1 tags. **f**, MYC-activated and MYC-repressed genes stratified by the ratio of MYC/MIZ1 bound to each promoter. *P* values, Chi-squared test. **g**, Expression of MYC-repressed genes. *x*-axis shows difference in expression between MYC(V394D) (MYC(VD)) and MYC (MYC(WT)); *y*-axis shows difference in MYC-dependent expression in MIZ1-depleted cells (shMIZ) relative to control (ctr) cells. Each bin has 83 genes. *P* value, two-tailed Student's *t*-test.

(H3K4me3; Fig. 3b)[1,2,15]. Genes bound by MIZ1 included genes transcribed by RNA Pol III (Extended Data Fig. 6c–f). MIZ1 was not bound at enhancers (Extended Data Fig. 6g)[2].

In primary cells, MIZ1 binds to a small number of sites, which contain the direct MIZ1-binding sequence[18]. This sequence was detectable in only 99 of the 1,000 promoters that were most strongly bound by MIZ1 and MYC, suggesting that MYC and MIZ1 can bind as a complex (Extended Data Fig. 7a). Consistently, a MYC-binding peak overlapped with a MIZ1 peak at 2,879 core promoters ($P < 2.2 \times 10^{-308}$ based on a hypergeometric distribution; Extended Data Fig. 7b). MIZ1-bound chromatin could be re-precipitated by anti-MIZ1, anti-MYC or anti-MAX antibodies, demonstrating that all three proteins co-occupy target promoters (Fig. 3c and Extended Data Fig. 7c). Binding of MIZ1 to promoters containing an E-box depended on MYC, whereas binding to promoters with a MIZ1-binding motif did not (Fig. 3d and Extended Data Fig. 7d). Induction of MYC recruited MIZ1 to promoters with E-boxes (Extended Data Fig. 7e; E-box, PPRC1; MIZ1-binding site, VPS72). Conversely, MYC and MAX occupied MIZ1 promoters in the absence of any E-box (Extended Data Fig. 7c, e, f). MYC and MIZ1 also bound to promoters lacking E-boxes and MIZ1-binding sites, albeit with lower occupancy, suggesting that MYC and MIZ1 can bind to target sites on chromatin independently of each other or co-operatively as a complex (Extended Data Fig. 7a).

Both MYC and MIZ1 are transcriptional activators that form a repressive complex upon binding to each other[19], suggesting that the MYC/MIZ1 ratio at each promoter affects the direction of the response to MYC. Supporting this hypothesis, the ratio of MYC to MIZ1 sequence tags at MYC-repressed promoters was around 1, whereas MYC-activated genes

showed a higher MYC/MIZ1 ratio (Fig. 3e, f). Genes with a lower MYC/MIZ1 ratio of sequence tags were regulated by MIZ1, but not MYC (Fig. 3f), and their promoters were enriched for the presence of a MIZ1-binding sequence (Extended Data Fig. 8a, b; $P = 0.00035$, Chi-squared test with Yates correction). MYC-repressed promoters had E-boxes (see Fig. 1), but lacked MIZ1-binding sequences, and were enriched for binding sites of SP1, which binds to both MYC and MIZ1, indicating that protein–protein interactions affect the MYC/MIZ1 ratio at each promoter (Extended Data Fig. 8c)[20,21]. shRNA-mediated depletion of MIZ1 had no effect on gene activation by MYC but de-repressed by twofold or more 259 of 680 genes that were repressed by MYC, demonstrating that MIZ1 is required for repression of a large fraction of, but not all, MYC target genes (Fig. 3g and Extended Data Fig. 8d). For all MYC-repressed genes, the extent of de-repression by shMIZ1 correlated with their enhanced expression in cells expressing MYC(V394D) instead of MYC; MYC (V394D) is compromised in binding to MIZ1 (Fig. 3g)[22]. Transcriptional activation by MYC(V394D) was unimpaired (Extended Data Fig. 8e).

To confirm that complex formation with MIZ1 is required for repression by MYC, we expressed doxycycline-inducible MYC or MYC(V394D) in mouse $Miz1^{flox/flox}$;CreER fibroblasts. In these cells, loxP sites flank exons encoding the POZ domain and activation of CreER generates $Miz1^{\Delta POZ/\Delta POZ}$ cells (Fig. 4a and Extended Data Fig. 9a)[23]. MYC bound to 10,437 promoters, 4,985 of which were shared with Miz1 (of 6,876 Miz1-bound promoters; Fig. 4b). Induction of MYC repressed 2,013 genes (fold change (FC) > 1.5). In the presence of MYC, deletion of the POZ domain significantly enhanced expression of 1,599 genes, 541 of which were repressed by MYC ($P = 5.1 \times 10^{-125}$). For all MYC-repressed genes,
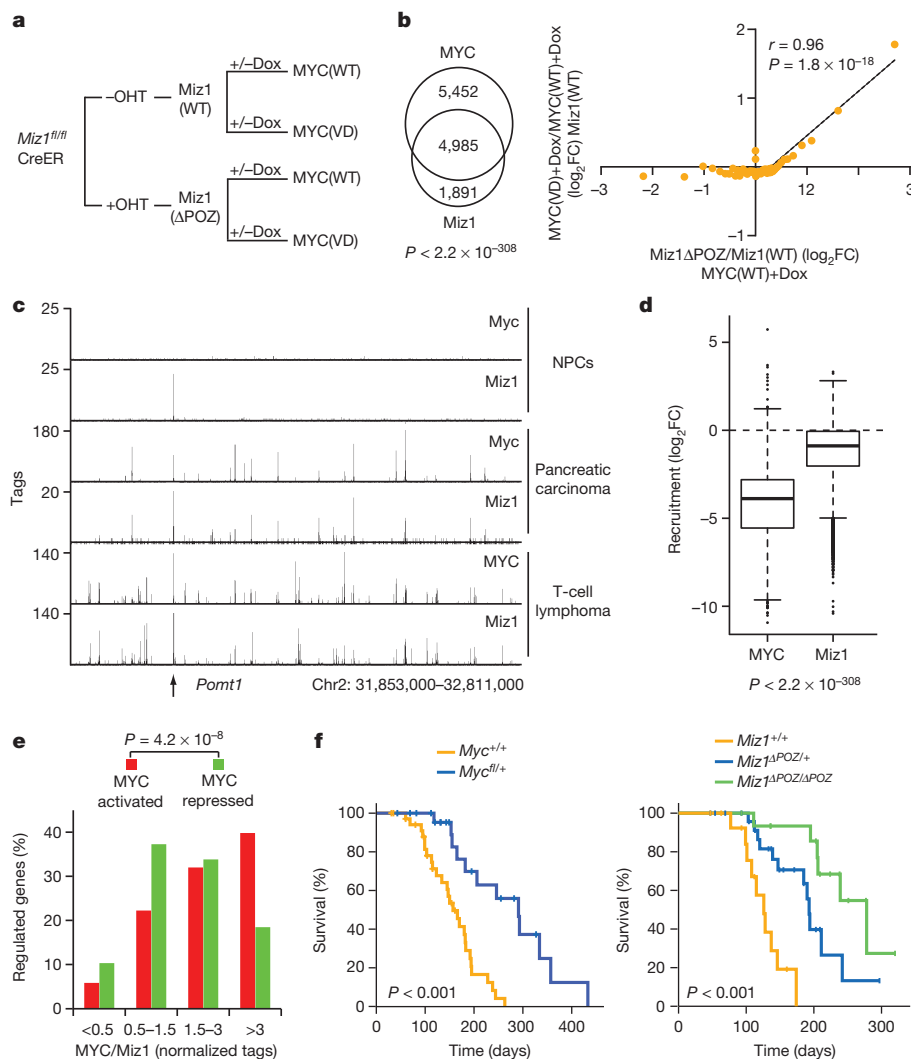
Figure 4 | Role of Miz1 in MYC-dependent gene regulation and tumorigenesis. a, Analysis of MYC-dependent changes in gene expression in mouse $Miz1^{flox/flox}$ fibroblasts. MYC(VD), MYC(V394D); WT, wild type. **b**, Left, MYC- and Miz1-binding sites. Right, x-axis shows change in expression upon activation of CreER after induction of MYC; y-axis shows difference in expression between MYC(V394D) and MYC. Each bin has 300 genes. **c**, ChIP-sequencing traces in neuronal progenitors (NPCs), T-cell lymphomas and pancreatic tumours. Pomt1 is a direct target of Miz1 (ref. 18). Chr2, chromosome 2. **d**, Boxplot of changes in binding at MYC/Miz1-binding sites in T-cell lymphomas upon repression of MYC. P values, one-sample t-test. **e**, MYC-activated and -repressed genes stratified by the ratio of MYC/Miz1 bound to each promoter. P values, Chi-squared test. **f**, Kaplan–Meier survival curves of $Pdx1$-$cre;LSL$-$Kras^{G12D/+};LSL$-$p53^{R172H/+}$ mice. Log-rank P values are shown.

the extent of de-repression upon addition of 4-hydroxy-tamoxifen (4-OHT) closely correlated with their enhanced expression in cells expressing MYC(V394D) instead of MYC (Fig. 4b). Deletion of the Miz1 POZ domain had virtually no effect on their expression in the absence of ectopic MYC (Extended Data Fig. 9b) and re-expression of MIZ1 partially restored MYC-dependent repression in $Miz1^{\Delta POZ/\Delta POZ}$ cells (Extended Data Fig. 9c). MYC(V394D) and MYC had no differential effects on expression of genes that require Miz1 for expression (Fig. 4b).

MYC bound to 36,497 sites in a doxycycline-regulatable mouse model of MYC-driven T-cell lymphoma, in which repression via Miz1 is required to suppress TGF-β-induced senescence[24]. Twelve-thousand one-hundred and twenty-five sites were shared with Miz1 (of 25,826 Miz1-binding sites; $P < 2.2 \times 10^{-308}$; Fig. 4c, Extended Data Fig. 9d and Supplementary Table 1). Doxycycline-mediated repression of MYC globally reduced chromatin binding of Miz1 (Fig. 4d). The decrease in Miz1 binding was not uniform, and joint promoters fell into two classes: at one, binding of Miz1 was unaffected by MYC, whereas at the other Miz1 binding was essentially eliminated when MYC expression was suppressed (Extended Data Fig. 9e). As in human cells, the direction of the transcriptional response to MYC closely correlated with the ratio of MYC/Miz1 bound to each promoter (Fig. 4e).

MYC and Miz1 co-occupied 1,214 promoters in tumours arising in $Pdx1\text{-}cre;LSL\text{-}Kras^{G12D/+};LSL\text{-}p53^{R172H/+}$ mice, in which oncogenic $Kras$ and mutant p53 are expressed in pancreatic progenitor cells (Fig. 4c and Extended Data Fig. 9f)[25]. Deletion of one allele of Myc in progenitor cells delayed tumorigenesis (Fig. 4f). Similarly, removal of one functional allele of $Miz1$ strongly delayed tumorigenesis, although $Miz1^{+/\Delta POZ}$ mice are phenotypically normal[18] (Fig. 4f). Tumorigenesis was further delayed upon removal of both functional $Miz1$ alleles.

We suggest that transcriptional amplification occurs when basal levels of MYC are low and association with MIZ1 is absent, and that changes in gene expression that occur specifically in response to oncogenic MYC levels can guide therapies targeting MYC-dependent tumours.

METHODS SUMMARY

U2OS cells were stably transfected with a doxycycline-inducible two-vector system (tet-on, Clontech) encoding a human MYC complementary DNA. Expression of MYC was induced with doxycycline ($1 \mu g \, ml^{-1}$) for 30 h. For depletion, HeLa cells were stably infected with a lentivirus containing an shRNA against MYC or control vector and cell pools were harvested after puromycin selection. ChIP and library generation were performed as described previously[3]. Antibodies are listed in Extended Data Fig. 4f. High-throughput sequencing was performed on an Illumina GAIIx platform following the manufacturer's instructions (see Methods for details). For re-ChIP, chromatin was eluted with the appropriate peptide and used for a second immunoprecipitation. Genome-wide expression analysis of HeLa cells was performed using an Agilent 44K Whole Human Genome Array. RNA-sequencing libraries were prepared with NEBNext RNA-Seq kit and sequenced as for the ChIP-sequencing libraries. Statistics and bioinformatic analyses of ChIP- and RNA-sequencing experiments and microarrays, as well as for all experiments, are described in Methods. Oligonucleotides used are described in Supplementary Table 2. The $Pdx1\text{-}cre;LSL\text{-}Kras^{G12D};LSL\text{-}Trp53^{R172H}$ mouse model has been described previously[25]. These mice were interbred with either $c\text{-}Myc^{flox}$ mice[26] or $Miz1^{flox}$ mice[23]. Mice on a mixed background were kept in conventional animal facilities and genotyped by Transnetyx. All procedures were performed under Home Office Licence and approved by the University of Glasgow Animal Welfare and Ethical Review Board. Mice were monitored daily and killed when they developed symptoms of late-stage pancreatic cancer. Tumour burden was assessed by gross pathology and histology.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Nie, Z. et al. c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. Cell **151**, 68–79 (2012).
2. Lin, C. Y. et al. Transcriptional amplification in tumor cells with Elevated c-Myc. Cell **151**, 56–67 (2012).
3. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell **133**, 1106–1117 (2008).
4. Zeller, K. I. et al. Global mapping of c-Myc binding sites and target gene networks in human B cells. Proc. Natl Acad. Sci. USA **103**, 17834–17839 (2006).
5. Horiuchi, D. et al. MYC pathway activation in triple-negative breast cancer is synthetic lethal with CDK inhibition. J. Exp. Med. **209**, 679–696 (2012).
6. Yustein, J. T. et al. Induction of ectopic Myc target gene JAG2 augments hypoxic growth and tumorigenesis in a human B-cell model. Proc. Natl Acad. Sci. USA **107**, 3534–3539 (2010).
7. Valentijn, L. J. et al. Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification. Proc. Natl Acad. Sci. USA **109**, 19190–19195 (2012).
8. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. Cell **141**, 432–445 (2010).
9. Wiese, K. E. et al. The role of MIZ-1 in MYC-dependent tumorigenesis. Cold Spring Harb. Perspect. Med. **3**, a014290 (2013).
10. Evan, G. I. et al. Induction of apoptosis in fibroblasts by c-myc protein. Cell **69**, 119–128 (1992).
11. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA **102**, 15545–15550 (2005).
12. Blackwell, T. K. et al. Binding of Myc proteins to canonical and noncanonical DNA sequences. Mol. Cell. Biol. **13**, 5216–5224 (1993).
13. Eberhardy, S. R. & Farnham, P. J. c-Myc mediates activation of the cad promoter via a post-RNA polymerase II recruitment mechanism. J. Biol. Chem. **276**, 48562–48571 (2001).
14. Eberhardy, S. R. & Farnham, P. J. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. J. Biol. Chem. **277**, 40156–40162 (2002).
15. Guccione, E. et al. Myc-binding-site recognition in the human genome is determined by chromatin context. Nature Cell Biol. **8**, 764–770 (2006).
16. Bouchard, C., Marquardt, J., Bras, A., Medema, R. H. & Eilers, M. Myc-induced proliferation and transformation require Akt-mediated phosphorylation of FoxO proteins. EMBO J. **23**, 2830–2840 (2004).
17. Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. **26**, 2119–2137 (2012).
18. Wolf, E. et al. Miz1 is required to maintain autophagic flux. Nature Commun. **4**, 2535 (2013).
19. Staller, P. et al. Repression of p15INK4b expression by Myc through association with Miz-1. Nature Cell Biol. **3**, 392–399 (2001).
20. Iraci, N. et al. A SP1/MIZ1/MYCN repression complex recruits HDAC1 at the TRKA and p75NTR promoters and affects neuroblastoma malignancy by inhibiting the cell response to NGF. Cancer Res. **71**, 404–412 (2011).
21. Gartel, A. L. et al. Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3. Proc. Natl Acad. Sci. USA **98**, 4510–4515 (2001).
22. Herold, S. et al. Negative regulation of the mammalian UV response by Myc through association with Miz-1. Mol. Cell **10**, 509–521 (2002).
23. Kosan, C. et al. Transcription factor Miz-1 is required to regulate interleukin-7 receptor signaling at early commitment stages of B cell differentiation. Immunity **33**, 917–928 (2010).
24. van Riggelen, J. et al. The interaction between Myc and Miz1 is required to antagonize TGFβ-dependent autocrine signaling during lymphoma formation and maintenance. Genes Dev. **24**, 1281–1294 (2010).
25. Hingorani, S. R. et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer Cell **7**, 469–483 (2005).
26. de Alboran, I. M. et al. Analysis of c-Myc function in normal cells via conditional gene-targeted mutation. Immunity **14**, 45–55 (2001).

Author Contributions S.W., F.L., E.W., K.E.W., B.v.E. and T.W. performed the experiments, S.H., H.D.-O. and M.T. characterized the MIZ1 637–807 antibody, S.W., F.L. and E.W. performed ChIP-sequencing experiments, J.M. analysed the pancreas model, F.L., F.R., M.B., M.F., S.W. and L.R. performed statistical analyses. E.W., L.Z., O.S., C.-L.W. and M.E. devised and supervised experiments. E.W. and M.E. wrote the paper and should be considered as senior authors of this study.

METHODS

Cell culture and shRNA-mediated gene silencing. HeLa, HEK293T and U2OS cells were obtained from ATCC. All cells were routinely tested for mycoplasma contamination. Cells were grown in DMEM (Sigma) supplemented with 10% fetal calf serum (Biochrome) and penicillin/streptomycin. U2OS cells were stably transfected with a doxycycline-inducible two-vector system MYC (tet-on; Clontec Laboratories) and selected with 2.5 µg ml^{-1} hygromycin (Invitrogen). Where indicated, MYC expression was induced by addition of doxycycline (30 h; 1 µg ml^{-1}). Lentiviruses expressing an shRNA against MYC were generated using a pLKO vector (TRC-consortium) by co-transfection with the packaging plasmid psPAX.2, and the envelope plasmid pMD2.G into HEK293 cells[27]. Cells were infected with lentiviral supernatants in the presence of 4 µg ml^{-1} polybrene (Sigma) for 24 h and selected with 2.5 µg ml^{-1} puromycin. Infected cells were harvested 72–96 h after infection for analysis. The long-lived protein degradation assay was performed as described[18].

Immunofluorescence. Induced and non-induced U2OS cells were washed twice with PBS and fixed with 3.7% paraformaldehyde. Cells were washed twice with 0.1 M glycine-PBS solution, permeabilized with 0.1% NP-40 and then blocked with 5% of fetal bovine serum. Anti-MYC antibody (N262) was diluted 1:200 in blocking solution and incubated for 45 min at 37 °C. After washing with PBS, Alexa-labelled anti-rabbit antibody incubation and PBS washing with Hoechst, the cells were mounted with Fluoromount (Sigma).

Flow cytometry analysis. Cells were labelled with 10 µM BrdU for 60 min. Afterwards, cells were fixed in 80% ethanol overnight at −20 °C, denatured in 2 M HCl with 0.5% Triton-X-100 for 30 min at room temperature, neutralized with 0.1 M sodium borate and stained with anti-BrdU FITC antibody. Finally, cells were resuspended in PBS with RNase A and propidium iodide, incubated for 30 min at 37 °C and analysed on a BD FACSCanto II flow cytometer. The forward light scatter (FSC) of unfixed cells was used as a relative measure of cell size.

ChIP and parallel sequencing. ChIP was performed as described previously[28]. Cells were treated with 1% formaldehyde for 10 min at 37 °C. After cell lysis, nuclei were re-suspended in RIPA buffer (10 mM Tris/HCl pH 7.5, 150 mM NaCl, 1% NP40, 1% deoxycholic acid (DOC), 0.1% SDS, 1 mM EDTA) and DNA was fragmented to a size <500 bp using a Branson sonifier. To obtain specific antibodies against MIZ1, a fragment of human MIZ1 (encompassing amino acids 637–803) was expressed in frame with a His-/GST-tandem tag in *Escherichia coli* (BL-21). Rabbits were immunized with the purified protein and antisera were tested for specificity by immunoblotting and immunoprecipitation. Antibodies were bound to Protein A/G-dynabeads or -sepharose (Invitrogen) and immunoprecipitated. Chromatin was eluted with 1% SDS or alternatively with 16 µg of the appropriate peptide (re-ChIPs, Abcam) and crosslinking was reverted overnight. For purification, the Qiagen PCR purification kit or chloroform/phenol extraction was used. ChIP-sequencing was performed as described before[3]. Purified DNA was end-repaired, A-tailed, ligated to Illumina adaptors, size-selected (200 bp) and purified with Qiagen gel extraction kit. DNA fragments were amplified by 18 cycles of PCR and library size was tested with the Biorad Experion system. The amount of library DNA was quantified using a picogreen assay and subjected to Illumina GAIIx sequencing according to the manufacturer's instructions. After base calling with the Genome Analyzer Data Collection Software, high quality PF-clusters (according to the CASAVA filter) were selected for further analyses.

qPCRs and global gene expression analysis. ChIP-DNA was analysed by qPCR in a MX3000P (Stratagene) with Sybr green Mix (Thermo). Immunoprecipitation or control 'cycle over threshold' (C_T) values were subtracted from the input C_T values and converted into per cent input. To measure mRNA amount, total RNA was extracted by using the RNAeasy reagent (Qiagen). PolyA$^+$-RNA was isolated with Sera-Mag Oligo(dT) Magnetic Particles (Thermo Scientific) and quantified with NanoDrop 1000. For expression analysis total cellular RNA was isolated with the RNAeasy reagent (Qiagen) including on-column DNA digestion (microarray analysis and RNA-sequencing) or extraction with peqGOLD TriFast (Peqlab) for RT–qPCR. For RT–qPCR first-strand was synthesized with M-MLV Reverse Transcriptase (Invitrogen) and random hexamer primers (Roche). Primer sequences are available in Supplementary Table 2. For genome-wide expression analysis in HeLa cells, cRNA was synthesized, labelled and amplified with a two colour Quick-Amp Labelling Kit (Agilent, p/n 5190/0444) according to the manufacturer's instructions. Labelled cRNA was hybridized on an Agilent 44K Whole Human Genome Array (G4845A 026652) and scanned with a Microarray scanner G2505C (Agilent) with 5 µm resolution at 20 bit. Raw data were generated using the Feature Extraction software v.10.1.1.1 from Agilent. The resulting intensity values for the red and green channels were normalized using the lowess method within the limma package in R/BioConductor. Regulated probes were selected on the basis that the logarithmic (base 2) average intensity value (A value) was ≥5. A threshold to indicate probes as differentially expressed between two samples (M value) was set at twofold change ($\log_2 FC = M$(treatment) − M(control)). Microarrays were carried out in technical duplicates for shMYC and in triplicates for empty vector.

For RNA-sequencing, PolyA$^+$-RNA was isolated from total with Sera-Mag Oligo-(dT) Magnetic Particles (Thermo Scientific). Library preparation was performed by using the NEBnext mRNA Library Prep Master Mix set for Illumina (E6100S/L) following the instruction manual. Briefly, PolyA$^+$-RNA was fragmented to generate 200-nucleotide fragments. First- and second-strand synthesis was performed and the resulting cDNA was end-repaired, ligated to NEBnext Adaptor, size-selected (250 bp) and purified with Qiagen gel extraction kit. cDNA was then amplified with 15 cycles of PCR and the resulting library was subjected to Illumina GAIIx sequencing according to the manufacturer's instructions. At several steps before and during the library preparation, the quality and integrity of RNA and cDNA was assessed through an Experion Automated Electrophoresis System (Bio-Rad).

ChIP-sequencing analysis. After quality control, the resulting fastq files were used for alignment to a precompiled hg19 reference index with BOWTIE v.0.12.7 (ref. 29). Unambiguously mapped reads were retained for subsequent generation of binding profiles, heat maps and calling of peaks. Downstream analyses were performed using R, Microsoft Excel or programs as described later. Mapped reads were used for peak calling with input or IgG sample as control and generation of .wig files with MACS v.1.4.2 (ref. 30) (the -keep-dup parameter was adjusted depending on the ChIP enrichment at the highest peaks), which were visualized using the Integrated Genome Browser software[31]. To create density distributions around TSSs or ChIP peaks and heat maps indicating co-occupancies, Seqminer v.1.3.3 (ref. 32) was used. Peak annotations were achieved using the 'closestBed' feature from the Bedtools suite v.2.11.2 (ref. 33) and the UCSC GoldenPath RefSeq database for human (hg19) or murine (mm9) genes. Intersections of ChIP-sequencing peaks were done with the 'intersectBed' tool from Bedtools and default parameters. MYC recruitment at promoters was calculated as the ratio of tags in a region ±100 bp around an annotated MYC peak summit after induction between the induced and non-induced situation. Recruitment of RNA polymerase was determined as the ratio of tags in a region from −100 to +300 bp relative to the TSS before and after MYC induction for all genes having an annotated MYC peak after induction. For recruitment of Pol II phosphorylated at Ser 2, a region from the transcription end site (TES) to +2 kb downstream was analysed. To avoid 0 tags in the uninduced situation, one tag was added to all regions. Pol II promoters were defined as regions −1.5 kb/+0.5 kb relative to the TSS.

RNA-sequencing analysis. Reads were alligned to the human genome with BOWTIE v.0.12.8. All further analysis were performed in R/BioConductor using BAM files. For differential expression analysis EdgeR was used[34].

Statistics and additional bioinformatic methods. Unless described differently in the figure legends, data are shown as means with standard deviation as error bars. For binned data plots, genes were sorted and grouped into equally sized bins, which were median-normalized if applicable. The median or mean value of each bin is shown and linear regression was used to illustrate trends. Fitting of the linear model to the data are given as Pearson's correlation coefficient (r) with corresponding P value using a Student's t-test. To test if two linear regressions differ significantly, an ANCOVA was applied. For the calculation of a significant overlap in a Venn diagram a hypergeometric distribution was used. Therefore the total number of base pairs covered by peaks in each ChIP-sequencing experiment was calculated as well as the number of base pairs in the overlapping region. As population size, all promoter regions (−1 kb to +0.5 kb relative to a TSS; total 60,336,000 bp (human) and 46,792,500 bp (mouse)) were used.

To test significant changes after Miz1 restoration in mouse $Miz1^{\Delta POZ/\Delta POZ}$ fibroblasts, a paired two-tailed t-test was applied. Changes in MYC and Miz1 occupancies after doxycycline-induced MYC-repression in T lymphocytes were determined to be statistically different from 0 using a one-sample t-test. To test significant changes in cell size, BrdU-positive cells, RNA amount and qPCR results, Student's t-tests were applied.

Functional analysis of gene groups was performed using DAVID[35]. Gene set enrichment analysis (GSEA) was performed with the C2 gene sets from the MSigDB (http://www.broadinstitute.org/gsea/msigdb). The number of permutations was set to 1,000 and significant ($P < 0.005$) gene sets related to MYC were selected. Binding motifs in MYC peaks within promoters of MYC-regulated genes were identified using the DREME algorithm within the MEME suite[36]. A region ±100 bp around the summit of a MYC peak was analysed.

For classification of *MYCN* amplified tumours, regulated genes from MYC-depleted HeLa cells (FC>1.5) and MYC-induced U2OS cells (FC>2), adjusted P value < 0.01) were converted into probe sets and duplicates were removed. These probe sets were used to compare to 38 data sets containing 19 *MYCN* amplified and 19 non-amplified samples by using a linear support vector machine (SVM) algorithm. The training set consisted of 96 samples. The accuracy gives the relative number of correct predictions and the Matthews correlation coefficient illustrates the fitting to the model.

Animal studies. Power analysis was used to calculate the minimum number of mice required in each cohort to enable us to detect a significant change in survival, based

on previous survival analyses in our laboratory. The *Pdx1-cre;LSL-Kras^{G12D};LSL-Trp53^{R172H}* mouse model has been described previously[25]. These mice were interbred with either *c-Myc^{flox}* mice[26] or *Miz1^{flox}* mice[23]. Mice on a mixed background were kept in conventional animal facilities and genotyped by Transnetyx. Mice were monitored daily and killed when they developed symptoms of late-stage pancreatic cancer. Tumour burden was assessed by gross pathology and histology. Numbers of animals were as follows: $Miz1^{+/+}$ ($n = 15$); $Miz1^{\Delta POZ/+}$ ($n = 27$); $Miz1^{\Delta POZ/\Delta POZ}$ ($n = 16$); $Myc^{+/+}$ ($n = 36$); $Myc^{fl/+}$ ($n = 25$). All animals were included in analyses; however, animals that needed to be euthanized owing to non-pancreatic disease (such as lymphoma or papillomas exceeding size limits) were censored and are shown on the Kaplan–Meier curves as cross-marks. Cohorts of mice were allocated to experimental groups based on genotype. Investigators responsible for monitoring and determining the endpoints of individual mice were blinded to their genotype. All procedures were performed under Home Office Licence and approved by the University of Glasgow Animal Welfare and Ethical Review Board.

27. Naldini, L., Blomer, U., Gage, F. H., Trono, D. & Verma, I. M. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. *Proc. Natl Acad. Sci. USA* **93,** 11382–11388 (1996).
28. Boyd, K. E. & Farnham, P. J. Myc versus USF: discrimination at the *cad* gene is determined by core promoter elements. *Mol. Cell. Biol.* **17,** 2529–2537 (1997).
29. Langmead, B. Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinformatics* **32,** 11.7.1–11.7.14 (2010).
30. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
31. Nicol, J. W., Helt, G. A., Blanchard, S. G., Jr, Raja, A. & Loraine, A. E. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25,** 2730–2731 (2009).
32. Ye, T. *et al.* seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39,** e35 (2011).
33. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
34. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
35. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).
36. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27,** 1696–1697 (2011).
37. Barski, A. *et al.* Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. *Nature Struct. Mol. Biol.* **17,** 629–634 (2010).
38. Grandori, C. *et al.* c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. *Nature Cell Biol.* **7,** 311–318 (2005).
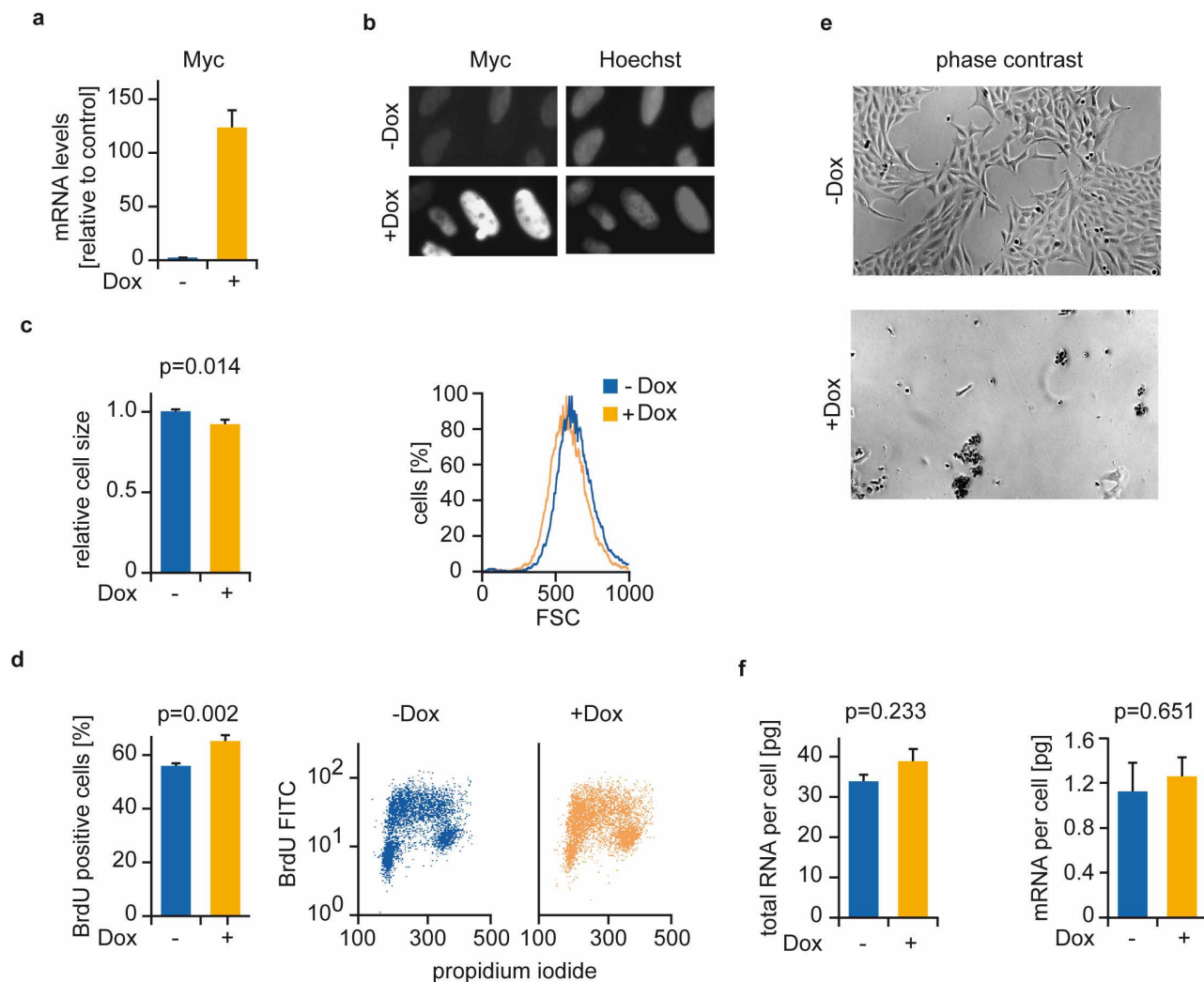
Extended Data Figure 1 | Characterization of U2OS cells expressing a doxycycline-inducible allele of *MYC*. a, RT–qPCR analyses demonstrating induction of *MYC* mRNA in response to doxycycline (Dox; 30 h; 1 µg ml^{-1}). Error bars show s.d. of triplicate technical assays. Control cells were treated with ethanol. **b,** Immunofluorescence using anti-MYC antibody and Hoechst staining of U2OS cells before and 30 h after induction of MYC. Original magnification: ×200. **c,** Cell size of U2OS cells before and 30 h after induction of MYC. Size was calculated from measurements of forward scatter (FSC) in arbitrary units. Errors bars show s.d. ($n = 3$; unless indicated otherwise,

n indicates the number of independent experiments in Extended Data Figs 2–10 legends). **d,** Percentage of BrdU-positive U2OS cells (left) and FACS analysis (right) documenting cell cycle distribution before and 30 h after induction of MYC with doxycycline ($n = 3$). **e,** Phase contrast pictures documenting induction of apoptosis by MYC. Pictures were taken 72 h after addition of doxycycline. Original magnification: ×50. **f,** Amount of total RNA (left) and of mRNA (right) per cell before and after induction of MYC. Error bars show standard error of the mean (s.e.m.) ($n = 4$).
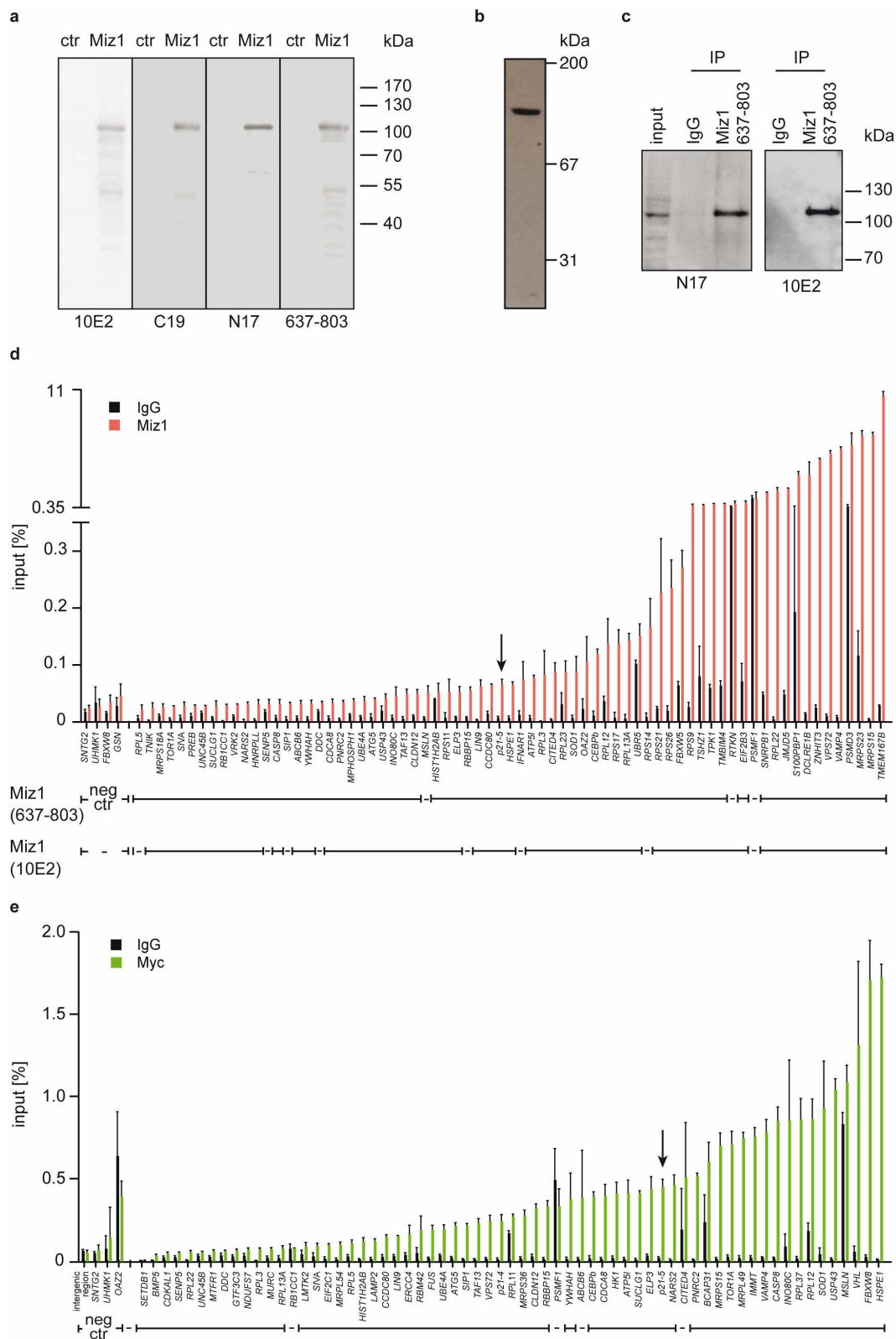
Extended Data Figure 2 | Controls for antibody specificity and validation of ChIP-sequencing data. a, HeLa cells were transfected with either control vectors or expression vectors encoding human MIZ1. Cell lysates were probed with the anti-MIZ1 antibodies used in this study. **b**, Immunoblot documenting specificity of the MIZ1 (637–803) antibody used for ChIP-sequencing experiments. The blot is from nuclear extracts of IMR90 human fibroblasts. **c**, Immunoprecipitations documenting that endogenous MIZ1 immunoprecipitated from HeLa cells by the MIZ1 (637–803) antibody is recognized by the N17 and 10E2 anti-MIZ1 antibodies. **d**, MIZ1 ChIP data for 71 randomly selected MIZ1-binding sites and 4 negative controls. Shown is the percentage of input DNA precipitated by the MIZ1 (637–803) antibody used for ChIP-sequencing and control IgG. Validation with a second MIZ1 antibody (10E2) is shown below. Minus sign indicates a non-significant enrichment over IgG ($q < 0.05$, heteroscedastic t-test). The arrow points to the start site of the *CDKN1A* gene ('p21-5'), a known target site of MIZ1. **e**, ChIP experiments analysing binding of MYC to randomly chosen promoters identified in the ChIP-sequencing analysis and four negative control sites. The experiment was performed using the antibody N262 in HeLa cells.
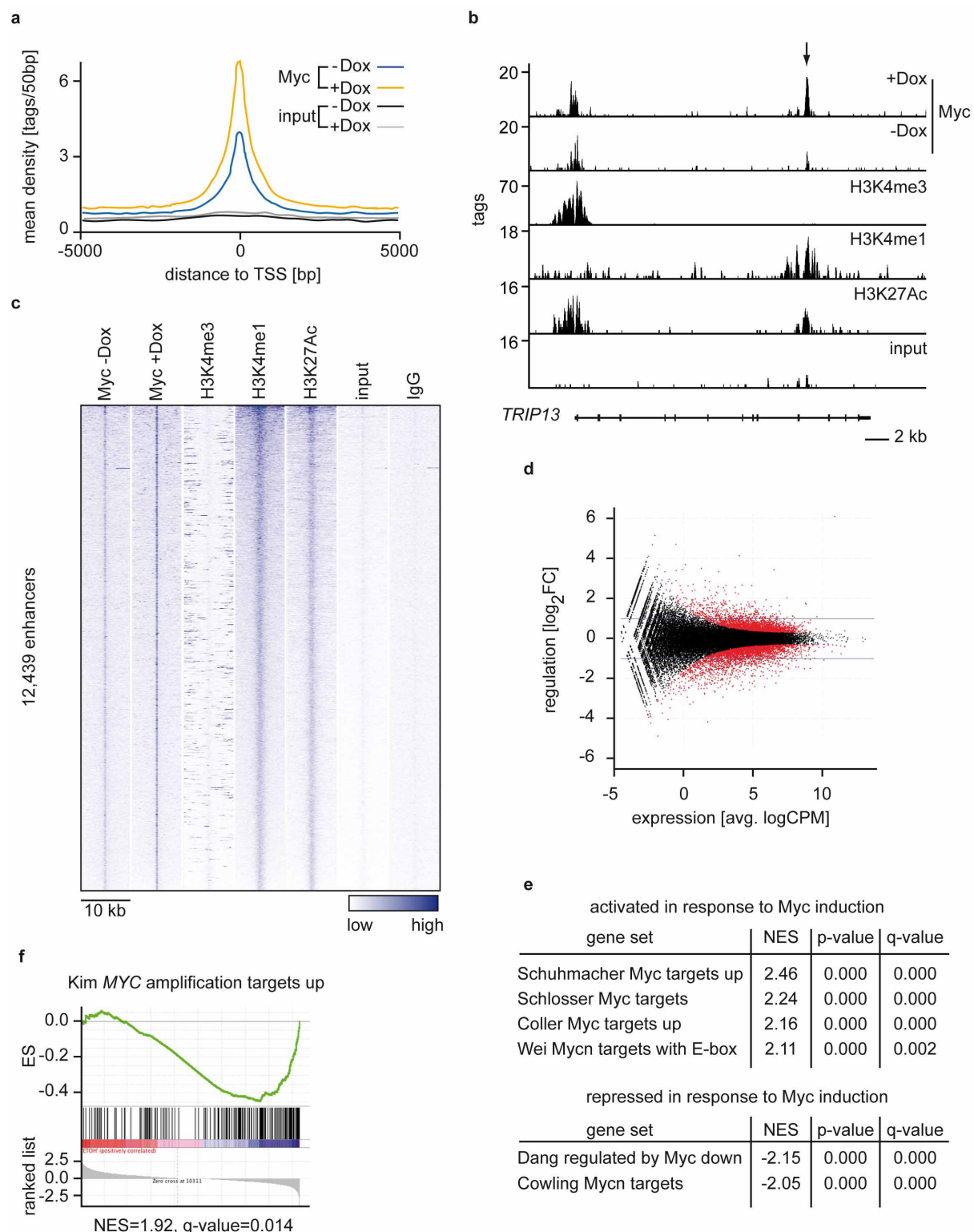
Extended Data Figure 3 | Binding of MYC to chromatin and MYC-dependent changes in gene expression in U2OS cells. a, Distribution of MYC tags around the TSS of all human Pol II genes with and without induction of exogenous *MYC*. **b**, Example of MYC binding to a promoter and an intragenic enhancer. Enhancers are identified by the presence of H3K4me1 and H3K27ac and the absence of H3K4me3. Exons are indicated as vertical bars, the untranslated region (UTR) shown as a thick black line. **c**, Heat map documenting binding of MYC to all enhancers identified in U2OS cells. Enhancer positions are centred according to MYC occupancy within a window of ±1 kb of the centre of the enhancer region and are sorted according to the number of H3K4me1 tags. **d**, The diagram shows the MYC-induced change in expression (plotted as log$_2$FC) versus total expression levels for all genes found in the RNA-sequencing as determined by RNA-sequencing. Red colour indicates significantly regulated genes ($q < 0.01$) ($n = 3$, biological replicates). CPM, counts per million. **e**, MYC-induced changes in gene expression are stereotypic. The panel shows a GSEA analysis of regulated genes in comparison to previously identified sets of MYC-regulated genes. **f**, Enrichment plot of a GSEA analysis showing that MYC-activated genes from U2OS cells identify lung tumours that have amplified *MYC*.

a

pLKO shMyc

Myc

Vinculin

b

p=0.019

relative cell size

pLKO shMyc

cells [%]

■ pLKO
■ shMyc

FSC

c

p=0.142

total RNA per cell [pg]

pLKO shMyc

d

repressed after Myc depletion

gene set	NES	p-value	q-value
Schuhmacher Myc targets up	-2.53	0.000	0.000
Schlosser Myc targets	-2.61	0.000	0.008
Dang Myc targets up	-2.05	0.000	0.008
Benporath Myc targets with E-box	-1.77	0.000	0.090

activated after Myc depletion

gene set	NES	p-value	q-value
Cowling Mycn targets	2.18	0.000	0.001
Dang regulated by Myc down	2.12	0.000	0.002
Coller Myc targets down	1.74	0.004	0.064

e

Kim *MYC* amplification targets up

ES

ranked list

shMyc (positively correlated)

Zero-cross at 10357

NES=2.06, q-value=0.009

f

	clone	method	supplier
Myc	N262	ChIP, immunofluorescence	Santa Cruz
	9E10	immunoblot	Santa Cruz
Miz1	aa637-803	ChIP, immunoblot	M. Teichmann
	10E2	ChIP, immunoblot	
	N17	ChIP, immunoblot	Santa Cruz
	C19	ChIP, immunoblot	Santa Cruz
Vinculin	hVIN-1	Immunoblot	Sigma-Aldrich
RNA Pol II	N20	ChIP	Santa Cruz
	8WG16	ChIP	Covance
pSer5 Pol II	4H8	ChIP	Covance
pSer2 Pol II	ab5095	ChIP	Abcam
BrdU-FITC	B44	FACS	BD Biosciences
Max	C17	ChIP	Santa Cruz
panAc H3	06-599	ChIP	Upstate (Millipore)
panAc H4	06-866	ChIP	Upstate (Millipore)
H3K4me1	ab8895	ChIP	Abcam
H3K4me3	ab8580	ChIP	Abcam
H3K27Ac	07-360	ChIP	Upstate (Millipore)

Extended Data Figure 4 | Oncogenic levels of MYC establish a tumour-cell-specific pattern of gene expression in HeLa cells. a, Immunoblots documenting MYC levels upon infection of HeLa cells with either control or lentiviruses expressing shRNA-targeting MYC. All analyses are shown for pools of stably infected and selected cells. **b**, Cell size of control and MYC-depleted HeLa cells. Size was calculated from measurements of forward scatter (FSC) in arbitrary units. Errors bars show s.d. (*n* = 3). **c**, Cellular RNA content of control and MYC-depleted HeLa cells (*n* = 3). **d**, The panels show GSEA analyses of gene sets regulated upon depletion of MYC in HeLa cells demonstrating a close overlap with previously identified sets of MYC- and MYCN-regulated genes. **e**, Stratification of human lung tumours by genes regulated in response to MYC depletion in HeLa cells. Shown is a GSEA enrichment plot with a gene set defining *MYC*-amplified lung tumours. **f**, Antibodies used in this study.
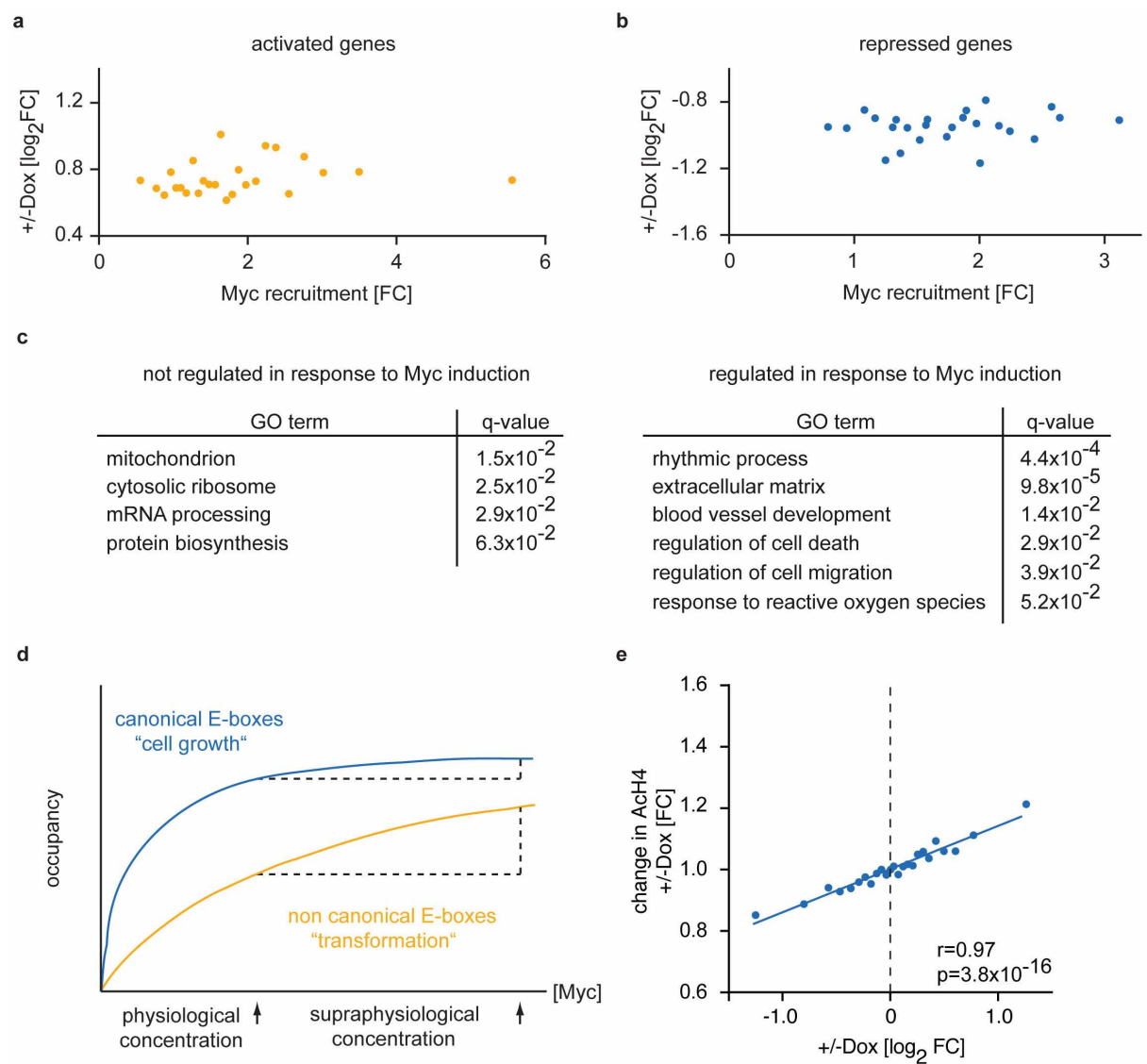
Extended Data Figure 5 | MYC recruitment at enhancer regions and gene regulation. **a**, **b**, Diagram depicting the change in gene expression versus the change in occupancy (plotted as fold recruitment) upon induction of MYC expression for 3,766 enhancers that are bound by MYC and the nearest genes in U2OS cells. Activated ($n = 1,457$) (**a**) and repressed genes ($n = 2,309$) (**b**) are shown in each diagram with 25 equally sized bins. **c**, GO terms of 500 MYC-bound genes that show the smallest change in gene expression and of the 500 most strongly MYC-regulated genes. **d**, Schematic drawing summarizing the results. The annotations 'cell growth' and 'transformation' are meant to summarize the GO terms. **e**, MYC-dependent changes in histone H4 acetylation. The diagram summarizes the change in histone H4 acetylation at the promoters of all MYC-bound genes in response to induction of MYC in U2OS cells. Data are stratified by the MYC-induced change in expression (plotted as $\log_2$FC). Each dot represents median values of a bin of 386 genes.
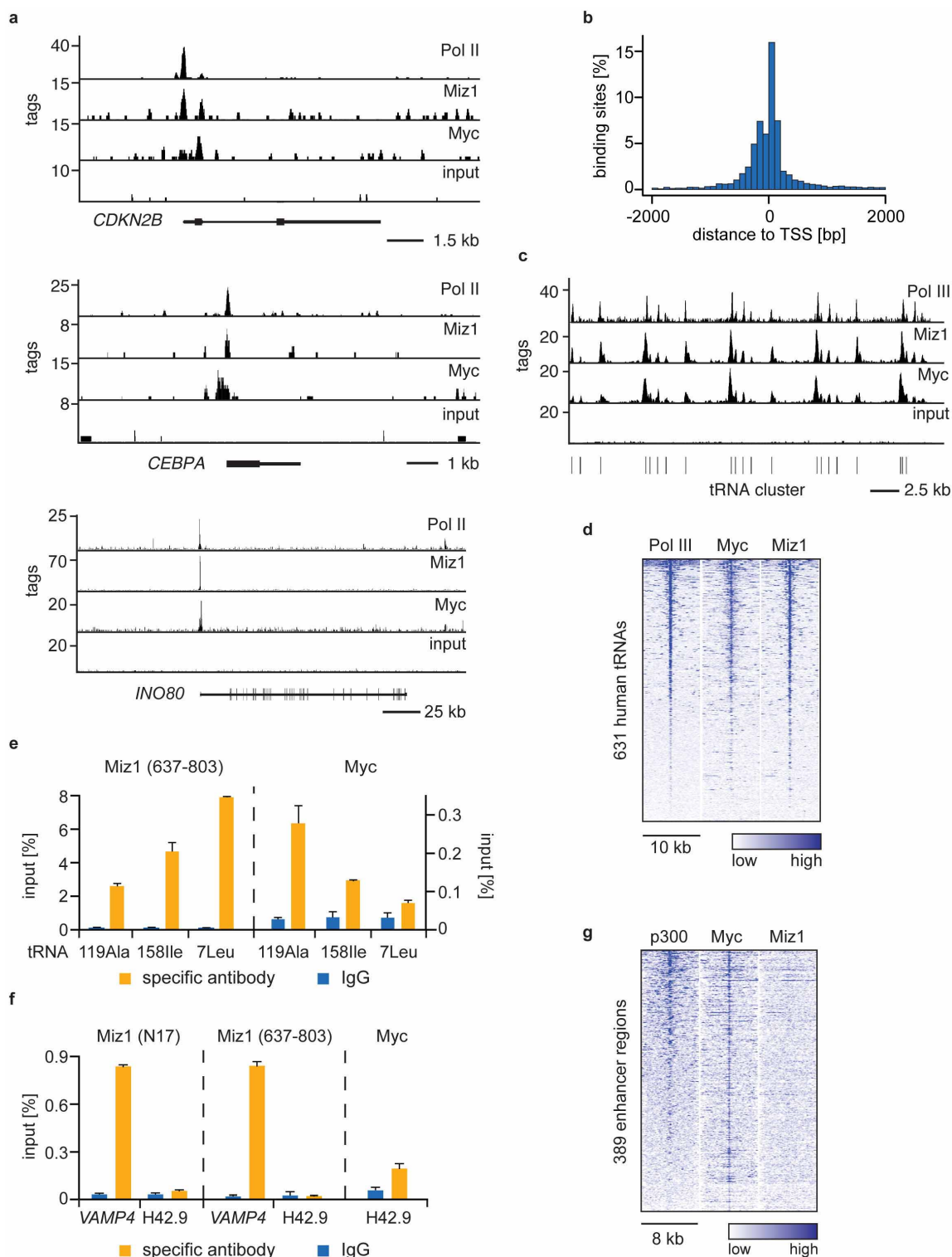
Extended Data Figure 6 | Additional characterization of MIZ1-binding sites on chromatin in HeLa cells. a, Examples of ChIP-sequencing traces for two known MYC/MIZ1 target genes (*CDKN2B*, *CEBPA*) and *INO80* (an E-box-containing promoter). **b**, Histogram showing the distribution of MIZ1-binding sites relative to TSSs. The positions of MIZ1 peaks are plotted relative to the next TSS of genes listed in the UCSC GoldenPath RefSeq database. **c**, Genome browser picture documenting binding of MYC and MIZ1 to a cluster of transfer (t)RNA genes (chr1:159,675,000–159,710,000 bp). ChIP-sequencing profiles for RNA Pol III binding are taken from ref. 37. Vertical lines indicate individual tRNA genes. **d**, Heat map documenting genome-wide occupancy of MYC, MIZ1 and RNA Pol III at tRNA loci. **e**, ChIP experiments documenting binding of MYC and MIZ1 to the indicated

human tRNA genes. The graphs show a representative result with error bars depicting s.d. of technical triplicates from one representative experiment (*n* = 3). **f**, ChIP experiments analysing binding of MYC and MIZ1 to ribosomal (r)DNA clusters indicated by the oligonucleotide 'H42.9'. The negative result for MIZ1 was confirmed using multiple primers to scan rDNA gene clusters, which previously has been shown to bind MYC[38]. Immunoprecipitations were performed in biological duplicates. The graphs show a representative result with error bars depicting s.d. of technical triplicates from one experiment. **g**, The heat map shows occupancy of 389 genomic regions predicted to be enhancers in HeLa cells. Enhancer regions were sorted according to p300 (GEO accession number GSM935500) occupancy and centred to the strongest MYC occupancy within a window of ±1 kb of the centre of the enhancer region.
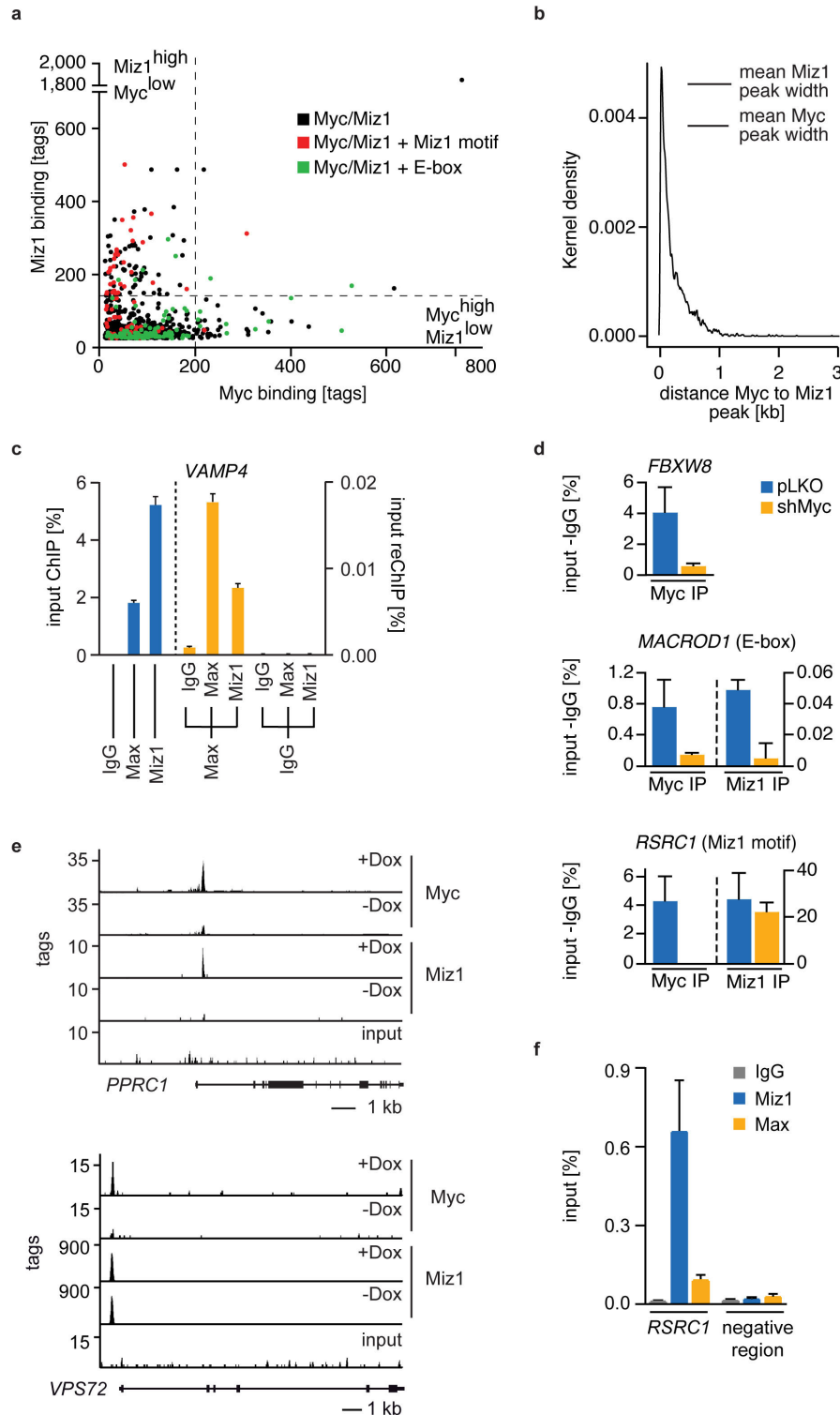
Extended Data Figure 7 | Characterization of two distinct classes of MYC/MIZ1-binding sites. a, The diagram shows the number of tags for MYC and MIZ1 for the 1,000 most strongly MIZ-bound core promoters transcribed by Pol II in HeLa cells. Red dots indicate binding sites that have a MIZ1 DNA-binding motif in the MIZ1 peak; green dots indicate binding sites that have an E-box (CACG/ATG) in the centre of the MYC peak (±60 bp). **b**, Gaussian Kernel distribution indicating the close relationship between MYC and MIZ1 peaks, with a bandwidth of 10 bp. Calculated was the distance between the summit of the MYC and the overlapping MIZ1 peak. The mean peak width (MYC, 687 bp; MIZ1,736 bp) is indicated by bars in the diagram. **c**, Re-ChIP experiment of MAX and MIZ1 at the *VAMP4* promoter. Left, primary ChIP; right, re-ChIP. IgG, control IgG. Error bars show s.d. of technical

triplicates (*n* = 2). **d**, ChIP experiments documenting binding of MIZ1 and MYC in control and MYC-depleted cells to the *MACROD1* (E-box) and *RSRC1* (MIZ1-binding motif) promoters. *FBXW8* was used as a control for MYC depletion from chromatin on a gene that was not MIZ1-bound. To control for experimental variation, values are normalized relative to the amount of Pol II precipitated in parallel reactions and background IgG signals were subtracted. **e**, Recruitment of MIZ1 by MYC to joint binding sites in U2OS cells. Representative browser pictures are from U2OS cells before and after induction of MYC. **f**, ChIP experiment documenting MAX and MIZ1 binding to the *RSRC1* (MIZ1-binding motif) promoter. The negative control region is on chromosome 11:79,556,864–79,556,942 bp.
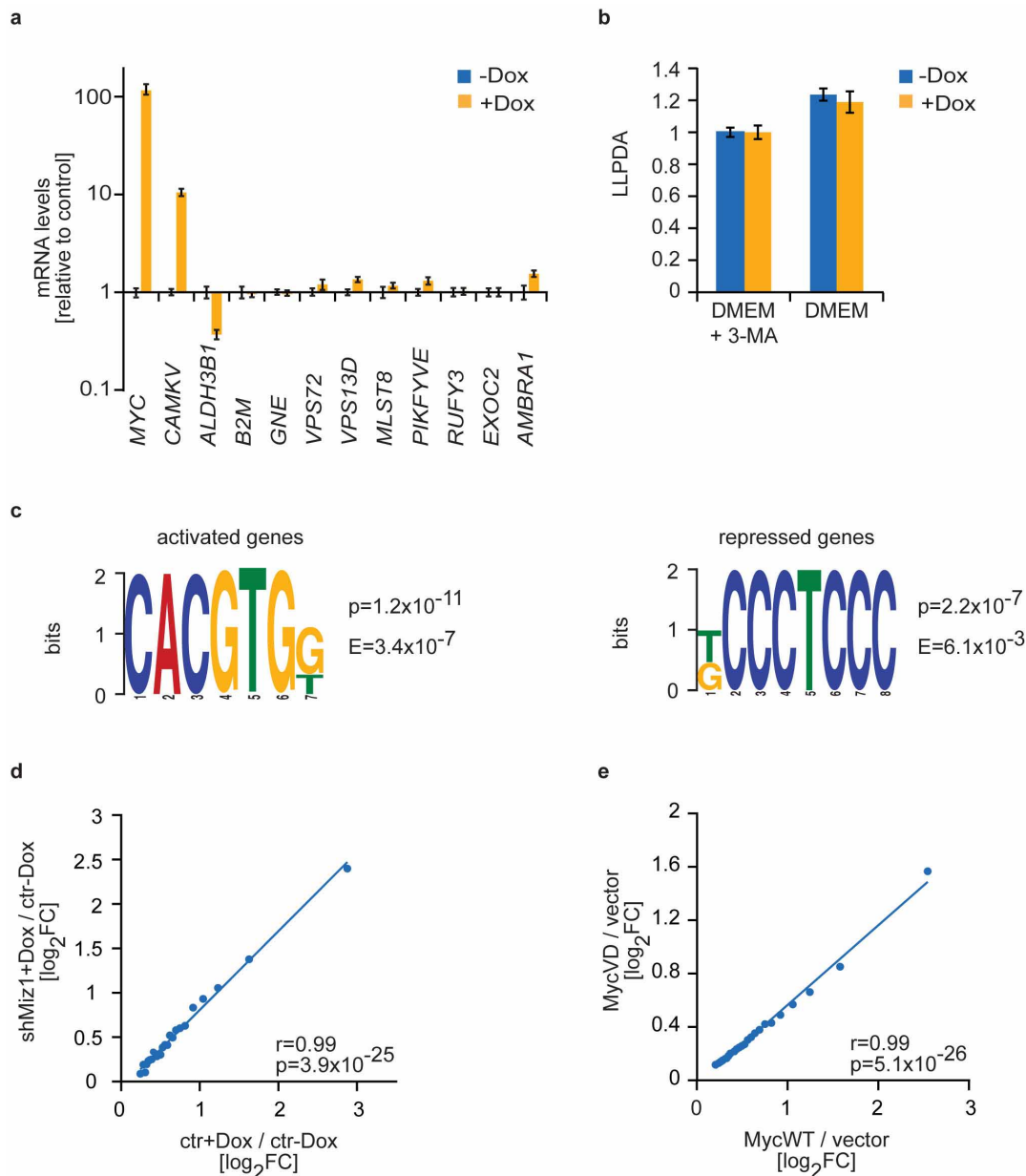
Extended Data Figure 8 | MIZ1 binding and transcriptional regulation by MYC. a, RT–qPCR assays showing no change in expression of genes that have a MIZ1 consensus binding sequence in their promoter in response to addition of doxycycline to U2OS cells expressing a doxycycline-inducible allele of *MYC*. *CAMKV* and *ALDH3B1* are MYC-regulated genes used as controls. **b,** Long-lived protein degradation assay in U2OS cells in the presence and absence of doxycycline. Where indicated, 3-methyladenin (3-MA) was added to inhibit autophagy. Values are normalized to 3-MA sample. Error bars show s.d. of biological replicates ($n = 3$). **c,** DREME analysis showing the most strongly enriched motifs in the 200 most MYC-activated and MYC-repressed genes. For the analysis, MYC peaks (±200 bp around the summit) that are located in promoters of MYC-regulated genes were used. **d,** Diagram depicting transcriptional activation of 2,787 genes upon induction of MYC in U2OS cells expressing control shRNA or shRNA targeting MIZ1 ($r = 0.99$; $P = 3.9 \times 10^{-25}$). Each bin summarizes expression of 112 genes. **e,** Transcriptional activation of 4,637 genes upon expression MYC(V394D) (MYC(VD)) or wild-type MYC (MYC(WT)) in U2OS cells ($r = 0.99$; $P = 5.1 \times 10^{-26}$). One-hundred and eighty-six genes are present in each bin.
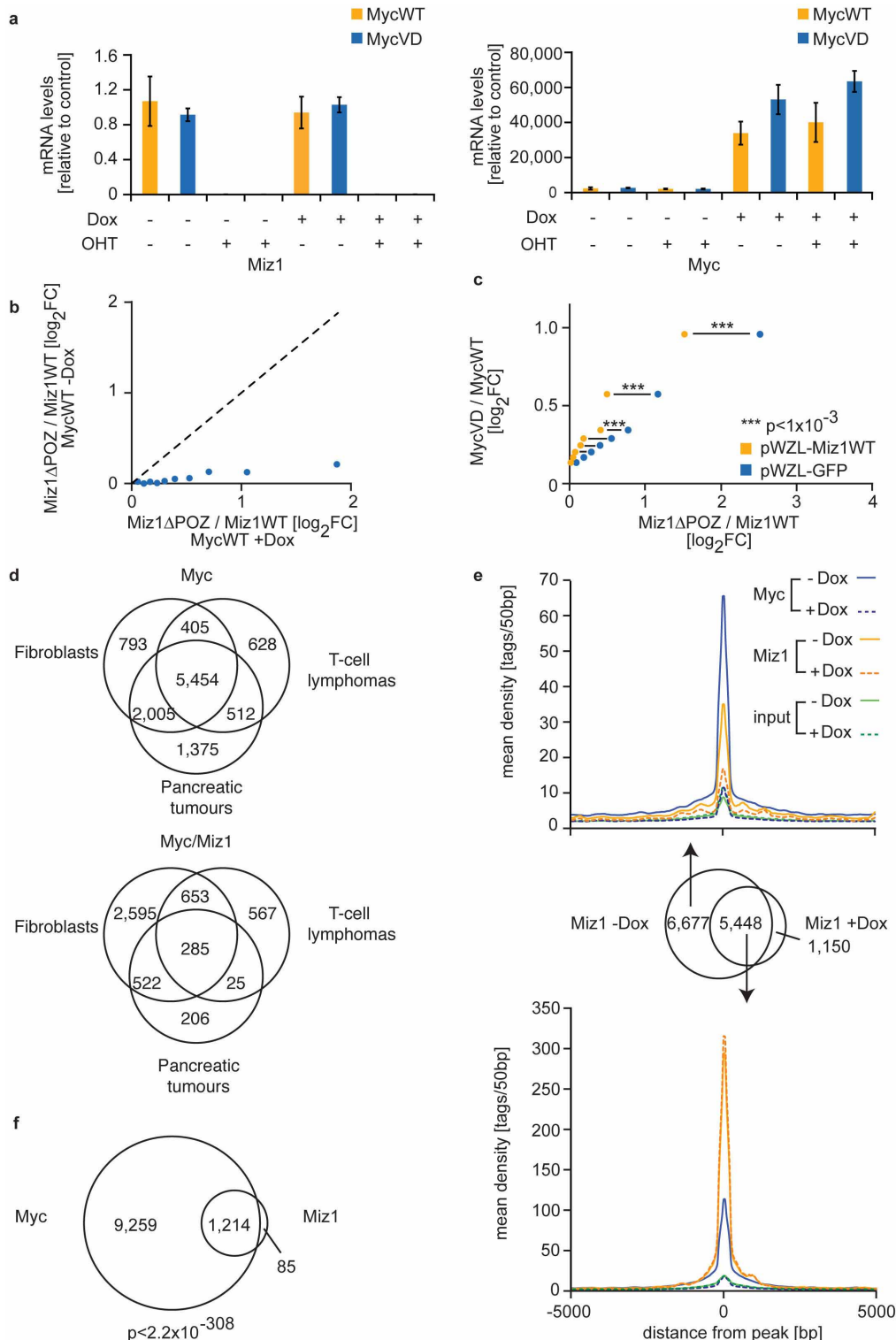
Extended Data Figure 9 | MIZ1 function in genetic models of MYC function. a, RT–qPCR assays showing the removal of exons encoding the POZ domain upon addition of 4-OHT and induction of MYC (Myc(WT)) or MYC(V394D) (Myc(VD)) in mouse *Miz1^flox/flox* fibroblasts expressing a CreER recombinase. **b**, Deletion of the Miz1 POZ domain has virtually no effect on expression of MYC-repressed genes in the absence of induced MYC. Data are shown for bins of genes that are upregulated by deletion of the POZ domain in the presence of MYC; x-axis shows 4-OHT-induced change in the presence of doxycycline, y-axis shows change in the absence of doxycycline. **c**, Restoration of wild-type MIZ1 in mouse *Miz1^ΔPOZ/ΔPOZ* fibroblasts re-establishes MYC-dependent repression. The data are plotted as in Fig. 4b. For each bin, a horizontal line indicates the change in expression upon restoration of MIZ1. *P* values indicate significance of alterations in gene expression for a given bin. *P* value, two-tailed paired Student's *t*-test. **d**, Venn diagrams showing overlap in Myc- and Myc/Miz1-binding sites in murine fibroblasts, T-cell lymphoma cells and pancreatic tumour cells. Target gene lists are shown in Supplementary Table 1. **e**, MYC and Miz1 binding to chromatin in T-cell lymphomas. The Venn diagram illustrates the change in *Miz1* DNA binding upon doxycycline-mediated removal of MYC. Of the 12,125 sites that are bound by MYC and Miz1 in T-cell lymphomas, Miz1 remains bound at 5,448 sites in the absence of MYC. **f**, Venn diagram showing binding of Myc and Miz1 to overlapping binding sites in cells established from pancreatic tumours arising in *Pdx1-cre;LSL-Kras^G12D/+;LSL-p53^R172H/+* mice ($P < 2.2 \times 10^{-308}$).

Extended Data Table 1 | Statistics of ChIP-sequencing reads used in the analysis

a

	HQ reads	mapped reads	binding sites	binding sites (FDR filter)	binding sites close to promoter	binding sites close to promoter (FDR 0.1)
Myc +Dox	8,481,707	7,947,678	45,696	45,645*	14,926	14,903
Myc -Dox	8,770,335	7,959,728	20,014	20,014*	8,401	8,401
Miz1 +Dox	7,187,304	5,636,366	28,186	28,186*	3,310	3,310
Miz1 -Dox	6,944,758	5,636,676	32,823	32,823*	4,195	4,195
Pol II +Dox	12,987,531	12,733,503	41,151	41,151*		
Pol II -Dox	12,966,907	12,736,047	37,695	37,695*		
Ser5 Pol II +Dox	18,083,944	17,558,426	93,397	93,397*		
Ser5 Pol II -Dox	17,898,479	17,563,122	73,426	73,426*		
Ser2 Pol II +Dox	8,035,906	7,526,547	71,081	71,073*		
Ser2 Pol II -Dox	7,806,811	7,544,065	62,951	62,951*		
H3K4me1 -Dox	7,264,234	7,128,027	59,256	49,074**		
H3K4me3 -Dox	6,334,300	6,138,231	23,592	23,013***		
H3K27Ac -Dox	3,921,915	3,860,049	40,188	39,613***		
panAc H3 +Dox	6,758,000	6,614,770				
panAc H3 -Dox	6,780,426	6,614,731				
panAc H4 +Dox	6,173,400	6,074,520				
panAc H4 -Dox	6,182,385	6,075,861				
input +Dox	20,666,229	20,348,902				
input -Dox	20,683,982	20,348,188				

b

	HQ reads	mapped reads	binding sites	binding sites (FDR 0.1)	binding sites close to promoter	binding sites close to promoter (FDR 0.1)
Myc	24,767,282	13,664,306	30,487	30,487	7,225	7,225
Miz1	19,125,018	10,786,542	6,124	6,086	3,286	3,270
input	19,787,036	10,278,874				

c

	HQ reads	mapped reads	binding sites	binding sites (FDR 0.1)	binding sites close to promoter	binding sites close to promoter (FDR 0.1)
Myc +Dox	14,453,559	13,687,224	708	20	43	2
Myc -Dox	16,598,890	13,685,221	36,948	36,947	7,752	7,752
Miz1 +Dox	14,375,816	12,935,144	19,252	12,669	5,779	4,686
Miz1 -Dox	13,612,330	12,936,309	28,627	25,826	4,867	4,572
IgG +Dox	11,461,671	10,741,676				
IgG -Dox	13,715,928	10,738,453				
input +Dox	12,435,449	11,585,216				
input -Dox	14,192,335	11,580,650				

d

	HQ reads	mapped reads	binding sites	binding sites (FDR 0.1)	binding sites close to promoter	binding sites close to promoter (FDR 0.1)
Myc	13,974,887	9,670,044	19,700	17,375	10,976	10,473
Miz1	13,912,438	9,675,548	2,625	2,625	1,299	1,299
IgG	13,798,404	9,671,602				

e

	HQ reads	mapped reads	binding sites	binding sites (FDR 0.1)	binding sites close to promoter	binding sites close to promoter (FDR 0.1)
Myc	15,767,147	10,029,924	27,910	27,910	10,437	10,437
Miz1	14,653,298	10,363,743	37,993	37,993	6,876	6,876
input	17,772,643	10,827,035				

a, ChIP-sequencing reads of U2OS cells used with peak calling. After performing peak calling using MACS and SICER software programs, peaks were filtered with different false discovery rate (FDR) values and used for subsequent analyses. Input control peaks were subtracted by MACS and SICER. Promoters are defined as −1.5 kb to +0.5 kb relative to the TSS.
b, ChIP-sequencing reads of HeLa cells used with peak calling. Promoters are defined as −1.0 kb to +0.5 kb relative to the TSS. **c**, ChIP-sequencing reads of T-lymphocyte cells and called binding sites using MACS software. Promoters are defined as −1.0 kb to +0.5 kb relative to the TSS. **d**, ChIP-sequencing reads of cells from pancreatic tumours arising in $Pdx1$-$cre;LSL$-$Kras^{G12D/+};LSL$-$p53^{R172H/+}$ mice and called binding sites. IgG control peaks were subtracted by MACS. Promoters are defined as −1.0 kb to +0.5 kb relative to the TSS. **e**, ChIP-sequencing reads in $Miz1^{\Delta POZ/\Delta POZ}$ fibroblasts and called binding sites using MACS software. Promoters are defined as −1.0 kb to +0.5 kb relative to the TSS.
* FDR = 0.1
** FDR = 1×10^{-4}
*** FDR = 0.05

LETTER

Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis

Arianna Sabò[1,2]*, Theresia R. Kress[1,2]*, Mattia Pelizzola[1], Stefano de Pretis[1], Marcin M. Gorski[2], Alessandra Tesi[1], Marco J. Morelli[1], Pranami Bora[1], Mirko Doni[2], Alessandro Verrecchia[2], Claudia Tonelli[2], Giovanni Fagà[2], Valerio Bianchi[1], Alberto Ronchi[1], Diana Low[3], Heiko Müller[1], Ernesto Guccione[3], Stefano Campaner[1] & Bruno Amati[1,2]

The c-*myc* proto-oncogene product, Myc, is a transcription factor that binds thousands of genomic loci[1]. Recent work suggested that rather than up- and downregulating selected groups of genes[1–3], Myc targets all active promoters and enhancers in the genome (a phenomenon termed 'invasion') and acts as a general amplifier of transcription[4,5]. However, the available data did not readily discriminate between direct and indirect effects of Myc on RNA biogenesis. We addressed this issue with genome-wide chromatin immunoprecipitation and RNA expression profiles during B-cell lymphomagenesis in mice, in cultured B cells and fibroblasts. Consistent with long-standing observations[6], we detected general increases in total RNA or messenger RNA copies per cell (hereby termed 'amplification')[4,5] when comparing actively proliferating cells with control quiescent cells: this was true whether cells were stimulated by mitogens (requiring endogenous Myc for a proliferative response)[7,8] or by deregulated, oncogenic Myc activity. RNA amplification and promoter/enhancer invasion by Myc were separable phenomena that could occur without one another. Moreover, whether or not associated with RNA amplification, Myc drove the differential expression of distinct subsets of target genes. Hence, although having the potential to interact with all active or poised regulatory elements in the genome[4,5,9–11], Myc does not directly act as a global transcriptional amplifier[4,5]. Instead, our results indicate that Myc activates and represses transcription of discrete gene sets, leading to changes in cellular state that can in turn feed back on global RNA production and turnover.

We first analysed the genomic distribution of Myc during B-cell lymphomagenesis *in vivo*. We generated ChIP-seq profiles in B cells from young non-transgenic (control, C) and Eµ-*myc* transgenic littermates (pre-tumour, P), and in lymphomas arising in adult Eµ-*myc* animals (tumour, T) (Extended Data Fig. 1a–j). Consistent with progressive increases in Myc mRNA and proteins levels, both binding intensity and the total number of binding sites progressively increased (around 7,000 sites in control, 17,000 in pre-tumour and 30,000 in tumour). Two-thirds of the Myc peaks in control were proximal to an annotated transcription start site (−2 to +1 kb from the transcription start site, henceforth referred to as 'promoter'). Although the numbers of proximal and distal peaks both increased in pre-tumour and tumour, most of the new binding sites were distal, with equal proportions of intra- and extragenic locations, and increasing distances from the nearest transcription start site. To chart active promoters and enhancers[12,13], we profiled RNA polymerase II (RNAPII) and the histone marks H3K4me3, H3K4me1 and H3K27ac: as expected, these features marked virtually all Myc-bound promoters (Fig. 1a and Extended Data Fig. 2a)[9,10]. Taking transcription start site annotations and H3K4me3 as references, Myc bound around 34% of active promoters in control, 66% in pre-tumour, and 87–94% in tumour samples. Instead, most unbound promoters showed no active histone marks or RNAPII (Fig. 1c). Distal Myc-binding sites bore H3K4me1, the activation mark H3K27ac and to

a lesser extent RNAPII, and showed the high H3K4me1/H3K4me3 ratios characteristic of enhancers (Fig. 1b, Extended Data Figs 2b, c and 3a). Out of 20 H3K4me1-positive regions characterized as active enhancers in mouse B cells[14], 2 were bound by Myc in control, 9 in pre-tumour and 18 in tumour (Extended Data Fig. 3b). Instead, distal enhancers with no Myc showed H3K4me1 but little or no H3K27ac or RNAPII (Fig. 1d), indicative of an inactive state. At both proximal and distal sites, the RNAPII and chromatin patterns in naive B cells (control sample) were similar to those in pre-tumour and tumour, preceding Myc at the same sites (Fig. 1a, b), and their intensities correlated with those of Myc binding (Extended Data Fig. 2d). In summary, Myc associated with regulatory elements that pre-existed in a poised or active state in naive B cells, the characteristic chromatin profiles of these sites anticipating Myc binding[9]. Most of these active elements were ultimately targeted in tumours, consistent with the concept of invasion[4].

We used RNA-seq to profile mRNA levels during tumour progression (Fig. 2a and Extended Data Fig. 4a). Normalizing to mean expression values yielded approximately 4,300 differentially expressed genes (DEGs) in pre-tumour and 3,900–4,600 in each tumour relative to control (Fig. 2b, c, Extended Data Fig. 4b and Supplementary Table 1). Most of the DEGs in pre-tumour were also DEGs in a least one tumour, with an equivalent amount of tumour-specific DEGs. Of all DEGs in lymphomas, 1,914 (27.8%) were common to the three samples (Fig. 2b). In all instances, more than half of DEGs showed increased expression, and approximately two-thirds of either class (up- or downregulated) had Myc bound to the promoter (Fig. 2c). Interpreting changes in mRNA levels is confounded by the fact that Myc can enhance total cellular RNA content[4,15], an effect that was confirmed in our pre-tumour and tumour samples (note the parallel increases in cell size[16], Fig. 2d, e). To account for this feature, we selected 754 mRNAs from our RNA-seq data and quantified them digitally with NanoString technology. This readily validated RNA-seq results, either as absolute or as differential expression relative to control (Extended Data Fig. 5a, b). Normalizing the NanoString counts per cell equivalents revealed upward shifts in the pre-tumour and tumour samples (Extended Data Fig. 5 c, d). As a consequence, very few down-regulated mRNAs were left, most mRNAs showing modest to strong upregulation of mRNA copies per cell (Fig. 2f). Importantly, this was observable whether or not Myc was bound to the promoter. Two implications follow from these data: first, rather than directly activating every promoter[4,5], Myc indirectly induced RNA amplification; second, direct regulatory cues (here the up- and downregulation of selected Myc-target genes) need to be discerned from global changes in RNA levels: this is achieved by normalizing RNA-seq profiles to mean expression, and not to cell equivalents[15].

The invasion of active regulatory elements by Myc and RNA amplification were described upon activation of a tet-*myc* transgene in the human B-cell line P493-6 (ref. 4), both of these effects were confirmed in our experiments (Extended Data Fig. 6a–c: 0 h, 1 h, 24 h, High).

[1]Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia (IIT), Via Adamello 16, 20139 Milan, Italy. [2]Department of Experimental Oncology, European Institute of Oncology (IEO), Via Adamello 16, 20139 Milan, Italy. [3]Institute of Molecular and Cell Biology, Singapore 138673, Singapore.
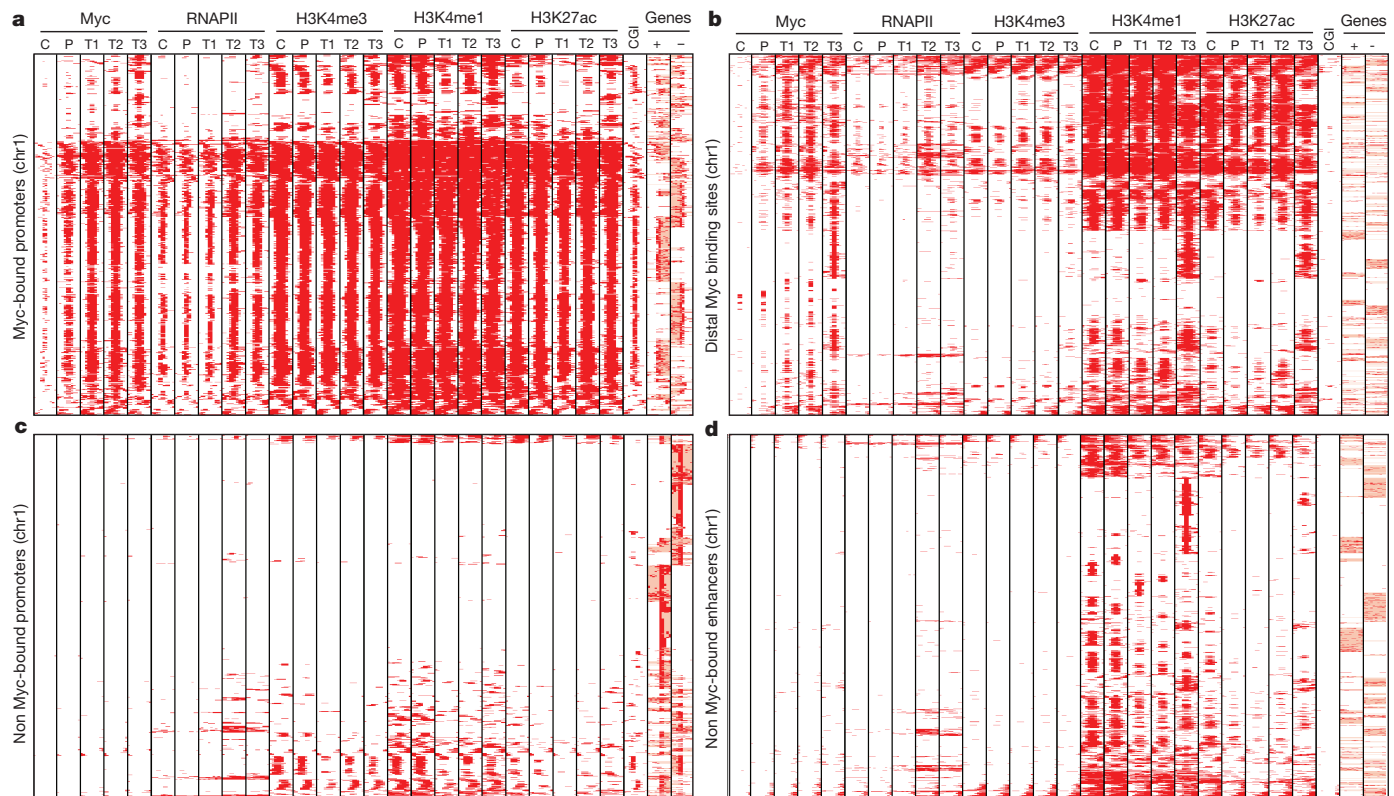*These authors contributed equally to this work.

Figure 1 | Increased Myc levels during lymphoma progression lead to invasion of accessible regulatory elements in the genome. **a**, The heatmap shows the distribution of Myc at annotated promoters. Each row represents a different genomic interval (6 kb width centred on Myc peaks). The panel includes every annotated promoter in chromosome 1 that was identified as Myc-associated by ChIP-seq in at least one of the experimental samples (C, P, T1, T2, T3). C, control; P, pre-tumour; T, tumour. For the same intervals, the distributions of RNAPII, H3K4me3, H3K4me1, H3K27ac, CpG islands (CGIs) and annotated genes (exons in red, introns in pink; + sense, − antisense strand) are also shown. **b–d**, As in **a**, for distal Myc binding sites (6 kb width centred on Myc peaks) (**b**), non Myc-bound promoters (6 kb width centred on the transcription start site) (**c**), and non Myc-bound enhancers, identified as distal H3K4me1-positive elements (6 kb width centred on H3K4me1 peaks) (**d**).

However, tet-repressed P493-6 cells (0 h) were quiescent, and Myc activation induced cell growth and proliferation[17] (Extended Data Fig. 6d, e). P493-6 cells can also proliferate, albeit more slowly, upon activation of a viral EBNA2–ER fusion protein, eliciting expression of endogenous Myc ('Low') at lower levels than with tet-Myc ('High')[18] and intermediate levels of RNA amplification (Extended Data Fig. 6a, f–h). Thus, RNA amplification correlated not only with Myc levels, but also with cellular activation and growth rates, precluding clear conclusions on cause-to-effect relationships. The same limit may apply to tumour cell lines proliferating with different Myc levels[4].

RNA amplification was also reported upon stimulation of B cells with lipopolysaccharide (LPS), and was attributed to a direct effect of endogenous Myc[5]. We repeated these experiments in B cells homozygous for a conditional knockout allele (c-myc[f/f]), allowing deletion of the gene before LPS treatment. Within the first 12 h, we confirmed a moderate increase in total RNA that was slightly reduced in the absence of Myc, as reported[5] (Extended Data Fig. 6i). This was accompanied by a similar increase in cell size, followed by a Myc-dependent proliferative response[7] (Extended Data Fig. 6j, k). However, both RNA content and cell size underwent striking increases at 24 h and 48 h, concomitant with plasma cell differentiation[6,19], which occurred even in the absence of Myc (Extended Data Fig. 6l)[20]. Thus, once again, neither of the above experiments allow us to discriminate between direct and indirect RNA amplification by Myc, the latter following from the up- and downregulation of specific gene sets in those cells (for example, ref. 21).

To further address this issue, we used mouse 3T9 fibroblasts expressing a conditional Myc-oestrogen receptor chimaera (3T9[MycER]). Endogenous Myc levels in exponentially growing 3T9[MycER] cells were sufficient to bind most of the active promoters (76%), as well as a sizeable number

of active enhancers (33%). These were invaded further upon MycER activation (95% and 59%, respectively), whereas inactive elements remained unbound (Extended Data Fig. 7a–c). Unlike the observations in B cells, MycER activation for up to 72 h caused no increase in cellular RNA content or cell size in fibroblasts (Fig. 3a, b). RNA-seq at 4, 8 and 16 h post-activation revealed differential regulation of 1,400 to 2,300 genes, of which up to 46% were downregulated (Fig. 3c and Supplementary Table 2). A time-course with quantitative reverse transcription PCR (qRT–PCR) on several genes showed that expression changes were rapid and at plateau before 8 h (Extended Data Fig. 7d). NanoString analysis validated the suppression and induction of selected mRNAs on a per-cell basis, without the general upregulation seen in B cells (Fig. 3d). Most importantly, these changes in gene expression were transcriptional in nature, as selective sequencing of newly synthesized RNA isolated through 4-thiouridine incorporation (4sU-seq) yielded a pattern highly consistent with RNA-seq (Fig. 3c). Hence, in either cell type, Myc contributed to direct transcriptional activation and repression of distinct groups of genes[1,2].

We then deleted the endogenous c-myc[f/f] gene in quiescent 3T9 fibroblasts, followed by serum stimulation. These cells showed a series of Myc-dependent effects, starting with the selective activation of approximately 300 Myc-dependent serum response (MDSR) genes in early G1, followed by S-phase entry[8] and by concomitant increases in either total RNA (back to the level observed in asynchronously growing cells: compare Fig. 3e and 3a) or mRNA molecules per cell as assayed by NanoString (Fig. 3f). Serum also induced a moderate increase in size that was attenuated in the absence of Myc (Fig. 3g). ChIP-seq analysis showed that endogenous Myc progressively binds a small subset of the sites bound by active MycER (Fig. 3h). Hence, the RNA 'amplification'
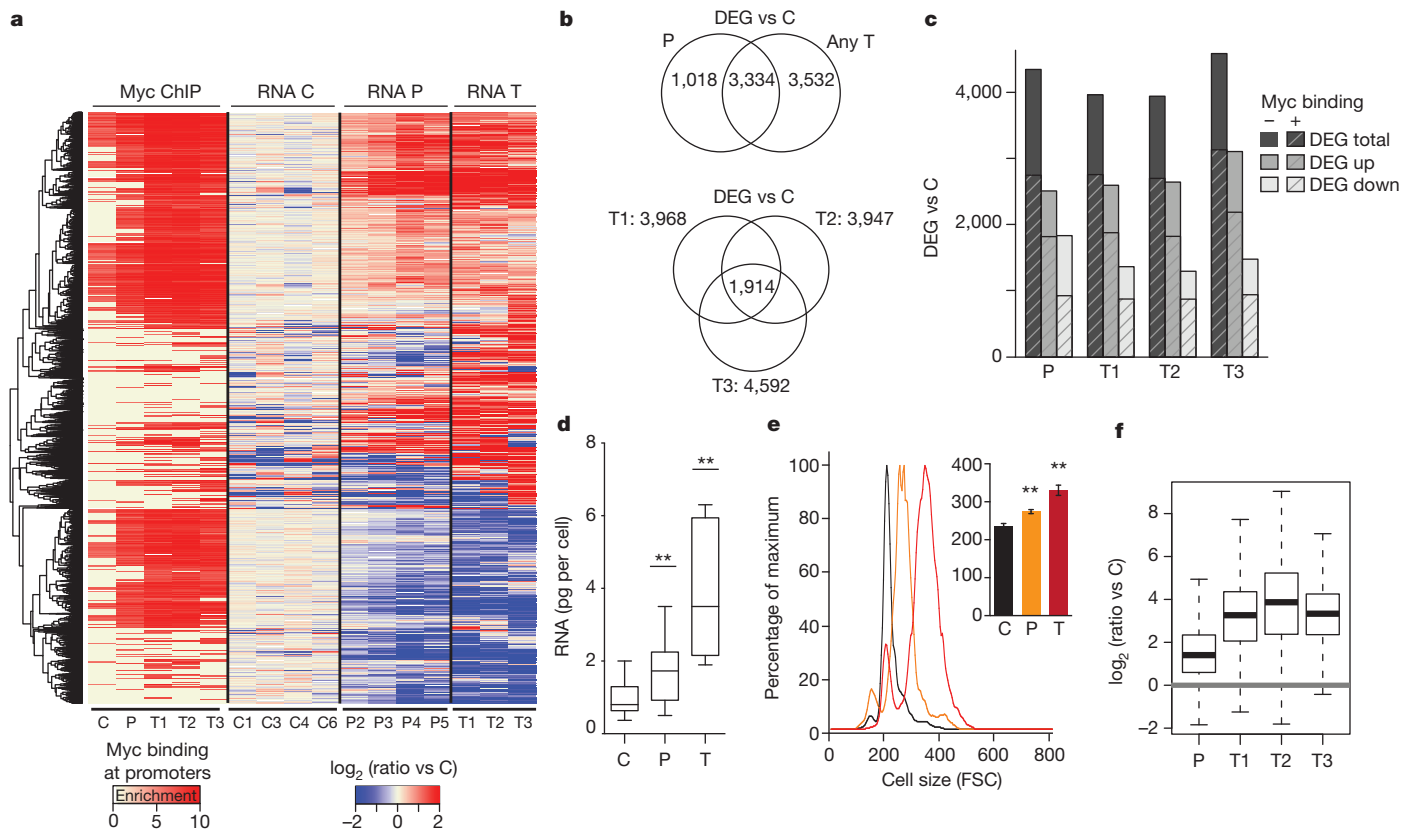
Figure 2 | Transcriptional amplification co-exists with selective up- and downregulation of specific Myc target genes. a, Heatmap of differentially expressed genes (DEGs) together with the indication about enrichment of Myc binding at the corresponding promoter. All DEGs in pre-tumour (P) or tumour (T) are relative to control (C). **b**, Venn diagrams showing the number of common DEGs between pre-tumour and tumour samples (top) and between the three tumours (bottom). **c**, Bar plot showing the number of DEGs (total, up and down) divided as Myc-bound and unbound, as indicated. **d**, Quantification of total RNA per cell ($n = 21$, 20 and 9 for control, pre-tumoral and tumoral mice, respectively; Student's t-test). **e**, Fluorescence-activated cell sorting (FACS) analysis of cell size distribution for representative control, pre-tumour and tumour samples (inset: mean ± s.d. of $n = 9$, 10 and 11 for control, pre-tumoral and tumoral mice, respectively; Student's t-test). **f**, Box plot showing gene expression changes in pre-tumour and tumour relative to control samples, as measured by NanoString with normalization to cell equivalents ($n = 3$ mice each). **$P < 0.001$.

effect seen in stimulated B cells[5] also occurs in fibroblasts, but without complete invasion of active regulatory elements.

Several aspects of our data address the nature of chromatin invasion by overexpressed Myc. First, the binding hierarchies among different promoters were conserved at different Myc levels in either B cells or fibroblasts (Extended Data Fig. 8a). Second, although all targeted promoters—even the weakest—were enriched for CpG islands and the H3K4me3 mark[8–10,22], only the strongest targets were enriched for consensus DNA binding motifs[8,9] (Extended Data Fig. 8b). The same was true at active enhancers, considering the H3K27ac mark. Third, MycER-bound sites in fibroblasts were hypersensitive to DNase I digestion, indicating higher accessibility (Extended Data Fig. 7c)[23]. Hence, invasion reflects increased interaction of overexpressed Myc with accessible chromatin domains, possibly favoured by recruitment (or tethering) by low-affinity protein–protein interactions[4,8,9,11], followed by non-specific engagement on DNA, and only at the strongest sites by sequence-specific DNA binding. The same order of events probably determines genome recognition by Myc or other transcription factors at physiological levels[11,23]. These observations imply that: (1) Myc cannot act as a pioneer factor[23] but instead requires prior opening of the targeted elements, as also observed in reprogramming experiments[10] and (2) invasion shall not a priori be equated with a productive engagement of Myc on all crosslinked elements[11].

Altogether, our results show that promoter/enhancer invasion and RNA amplification are separable and functionally unrelated consequences of Myc overexpression. Thus, rather than the direct upregulation of every active gene (Fig. 4a)[4,5], RNA amplification, when it occurs, is an indirect

effect of Myc, which acts primarily through the differential regulation of specific groups of genes (Fig. 4b). Among the functional categories enriched by Myc-regulated genes, we find a variety of DNA- and RNA-associated processes (Extended Data Fig. 9) that may be instrumental for RNA amplification either co- or post-transcriptionally, as exemplified by the Myc-repressed RNA-stabilizing factor ZFP36 (also known as TTP)[24]. Other cellular processes controlled by Myc and its target genes, such as nucleotide and energy metabolism, mitochondrial biomass, ribosome biogenesis or cell growth/size[3,16,17,25–27] may feed back on general transcriptional activity[28,29] (Fig. 4b). These feedback mechanisms are highly conserved, pre-date Myc in evolution and rely in part on the global regulation of transcriptional elongation (or stalling), a phenomenon that was also directly attributed to Myc[4,30]. Analysis of RNAPII distribution along Myc-induced genes in lymphomas and fibroblasts revealed composite increases in promoters and gene bodies, albeit without consistent effects on stalling indexes (Extended Data Fig. 10), warranting further discrimination between direct and indirect effects of this pervasive transcription factor. An essential corollary of our work is that the biology of Myc in physiology and disease must still be understood through comprehensive mapping of its target genes, as achieved here during B-cell lymphomagenesis. Our data constitute a unique resource for the functional characterization of these genes in tumour progression and maintenance.

Note added in proof: Cunningham *et al.* (*Cell* **157**, 1088–1103, 2014) recently reported a mechanism that could underlie indirect transcriptional amplification via a metabolic increase in total RNA production in the same Myc-transgenic model used here: in pre-tumoral Eμ-*myc* B
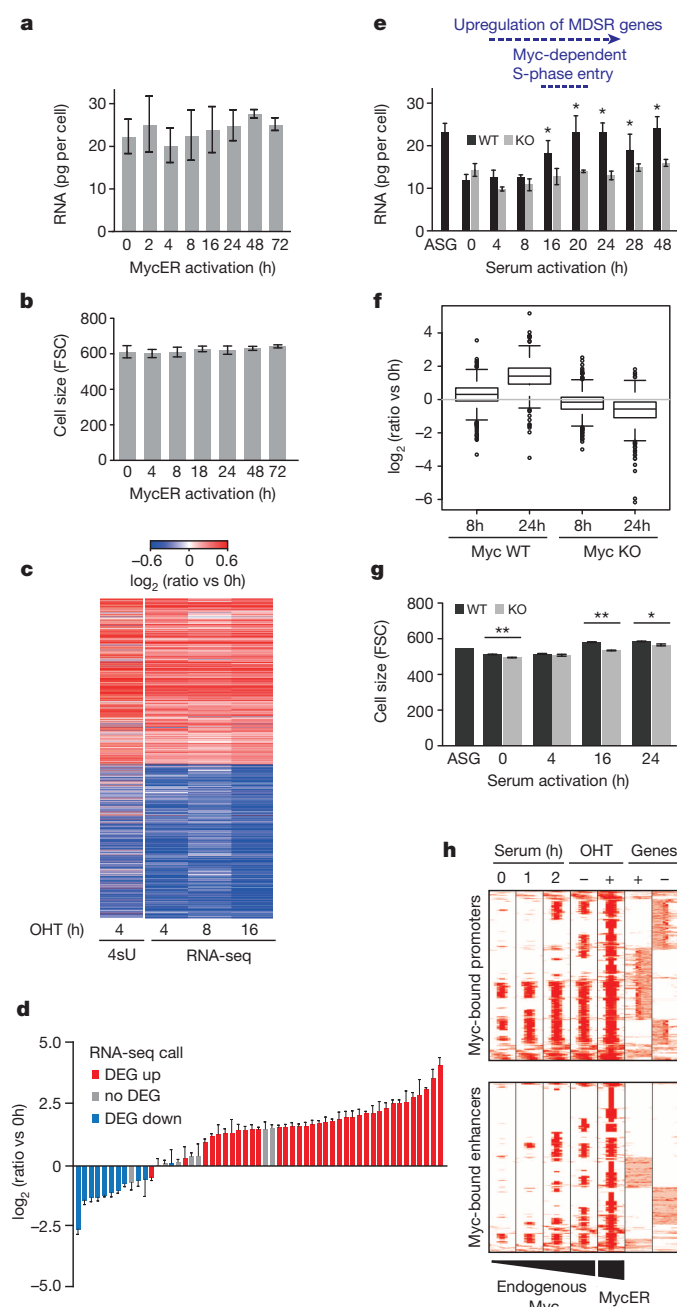
Figure 3 | RNA amplification and chromatin invasion are separable events in murine fibroblasts. a, Quantification of total RNA per cell (mean ± s.d.; $n = 3$). **b**, Cell size measurement by forward scatter (FSC, as mean ± s.d.; $n = 3$) after treatment of 3T9^{MycER} cells with 4-hydroxytamoxifen (OHT). **c**, Relative levels of nascent RNA (4sU-seq, single sample) and total mRNA (RNA-seq, $n = 3$) for up- and downregulated genes following MycER activation. **d**, NanoString validation of differentially expressed genes (DEGs) 8 h after OHT treatment (mean ± s.d.; $n = 3$; see gene list, see Methods). **e**, Quantification of total RNA per cell of asynchronously growing (ASG) or serum starved and released fibroblasts (mean ± s.d.; $n = 3$, Student's t-test, P values of induced total cellular RNA levels were calculated relative to 0 h). KO indicates Myc deletion before serum stimulation. As indicated on top, upregulation of Myc-dependent serum-response (MDSR) genes is observed from 4 h onwards, and Myc-dependent S-phase entry at approximately 16 h[8]. **f**, Box plot showing gene expression changes 8 h and 24 h after serum stimulation relative to starved cells (0 h), as measured by NanoString with normalization to cell equivalents (single sample). **g**, Cell size after serum stimulation. Values are FSC mean ± s.d.; $n = 3$; Student's t-test. **h**, Distribution of Myc binding at Myc-bound promoters and enhancers (as in Fig. 1). *$P < 0.05$, **$P < 0.001$.
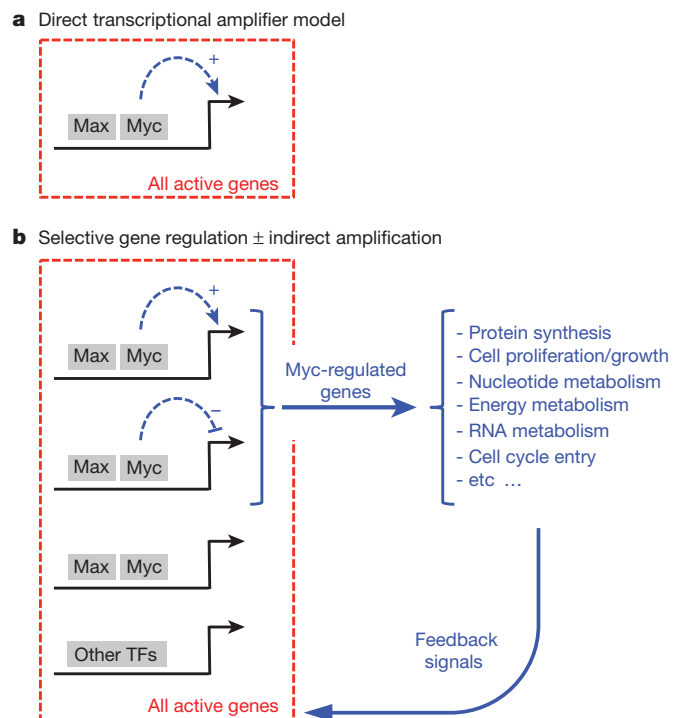
Figure 4 | An alternative mechanism for Myc-induced RNA amplification. a, Direct transcriptional amplification[4,5]. **b**, Indirect RNA amplification. Selective gene regulation by Myc leads to a variety of changes in cellular state, which in the appropriate physiological settings feed back on general RNA biogenesis and turnover (see main text). This scenario predicts that RNA amplification may happen in the absence of enhancer or promoter invasion by Myc and, conversely, invasion shall not systematically be associated with RNA amplification, as verified here.

cells, induction of the translation factor eIF4E led to augmented production of the nucleotide biosynthetic enzyme PRPS2, which in turn was required for the observed increase in RNA synthesis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Eilers, M. & Eisenman, R. N. Myc's broad reach. *Genes Dev.* **22**, 2755–2766 (2008).
2. Herkert, B. & Eilers, M. Transcriptional repression: the dark side of Myc. *Genes Cancer* **1**, 580–586 (2010).
3. Dang, C. V. MYC, metabolism, cell growth, and tumorigenesis. *Cold Spring Harbor Perspect. Med.* **3**, http://dx.doi.org/10.1101/cshperspect.a014217 (2013).
4. Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).
5. Nie, Z. *et al.* c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* **151**, 68–79 (2012).
6. Darzynkiewicz, Z., Traganos, F. & Melamed, M. R. New cell cycle compartments identified by multiparameter flow cytometry. *Cytometry* **1**, 98–108 (1980).
7. de Alboran, I. M. *et al.* Analysis of C-MYC function in normal cells via conditional gene-targeted mutation. *Immunity* **14**, 45–55 (2001).
8. Perna, D. *et al.* Genome-wide mapping of Myc binding and gene regulation in serum-stimulated fibroblasts. *Oncogene* **31**, 1695–1709 (2012).
9. Guccione, E. *et al.* Myc-binding-site recognition in the human genome is determined by chromatin context. *Nature Cell Biol.* **8**, 764–770 (2006).
10. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
11. Sabo, A. & Amati, B. Genome recognition by MYC. *Cold Spring Harbor Perspect. Med.* **4**, http://dx.doi.org/10.1101/cshperspect.a014191 (2014).
12. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nature Rev. Genet.* **12**, 7–18 (2011).
13. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
14. Lin, Y. C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nature Immunol.* **11**, 635–643 (2010).

15. Lovén, J. *et al.* Revisiting global gene expression analysis. *Cell* **151,** 476–482 (2012).
16. Iritani, B. M. & Eisenman, R. N. c-Myc enhances protein synthesis and cell size during B lymphocyte development. *Proc. Natl Acad. Sci. USA* **96,** 13180–13185 (1999).
17. Schuhmacher, M. *et al.* Control of cell growth by c-Myc in the absence of cell division. *Curr. Biol.* **9,** 1255–1258 (1999).
18. Yustein, J. T. *et al.* Induction of ectopic Myc target gene JAG2 augments hypoxic growth and tumorigenesis in a human B-cell model. *Proc. Natl Acad. Sci. USA* **107,** 3534–3539 (2010).
19. Kouzine, F. *et al.* Global regulation of promoter melting in naive lymphocytes. *Cell* **153,** 988–999 (2013).
20. Murn, J. *et al.* A Myc-regulated transcriptional network controls B-cell fate in response to BCR triggering. *BMC Genomics* **10,** 323 (2009).
21. Fan, J. *et al.* Time-dependent c-Myc transactomes mapped by array-based nuclear run-on reveal transcriptional modules in human B cells. *PLoS ONE* **5,** e9691 (2010).
22. Fernandez, P. C. *et al.* Genomic targets of the human c-Myc protein. *Genes Dev.* **17,** 1115–1129 (2003).
23. Guertin, M. J. & Lis, J. T. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr. Opin. Genet. Dev.* **23,** 116–123 (2013).
24. Rounbehler, R. J. *et al.* Tristetraprolin impairs Myc-induced lymphoma and abolishes the malignant state. *Cell* **150,** 563–574 (2012).
25. Graves, J. A. *et al.* Mitochondrial structure, function and dynamics are temporally controlled by c-Myc. *PLoS ONE* **7,** e37699 (2012).
26. Wang, R. *et al.* The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* **35,** 871–882 (2011).
27. Liu, Y. C. *et al.* Global regulation of nucleotide biosynthetic genes by c-Myc. *PLoS ONE* **3,** e2722 (2008).
28. Marguerat, S. & Bahler, J. Coordinating genome expression with cell size. *Trends Genet.* **28,** 560–565 (2012).
29. das Neves, R. P. *et al.* Connecting variability in global transcription rate to mitochondrial variability. *PLoS Biol.* **8,** e1000560 (2010).
30. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141,** 432–445 (2010).

Supplementary Information is available in the online version of the paper.

METHODS

Primary mouse B cells and cell lines. C57BL/6 Eμ-*myc* transgenic mice[31] were monitored twice a week for lymphoma development by peripheral lymph node palpation[32]. In all experiments, we used gender- and age-matched mice (both males and females) without randomization or blinding. Tumours and lymphoid organs were dissected and processed for molecular analysis as described[33]. Briefly, for control (C) and pre-tumour (P) samples, spleens of 6–8-week-old mice with no infiltration of peripheral lymph nodes were used. We obtained single-cell suspensions by pressing the spleen through nylon cell strainers and subsequent hypotonic lysis of red blood cells. To isolate B cells, we incubated single-cell suspensions with B220 MicroBeads (Miltenyi Biotech) and enriched them by magnetic cell sorting (MACS), according to the manufacturer's instructions (Miltenyi Biotech). Lymphoma samples (or tumours, T) composed primarily of tumour cells, were dissected from infiltrated lymph nodes and not purified further. At this stage, all samples were fixed directly for chromatin immunoprecipitation (see below) without any *in vitro* treatment or culture. Experiments involving animals were performed in accordance with the Italian laws (D.L.vo 116/92 and following additions), which enforce EU 86/609 directive (Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes).

P493-6 cells[34] were cultured in RPMI medium with 10% tetracycline-free serum, 2 mM L-Gln 1% penicillin/streptomycin and non-essential amino acids (NEAA) with either: (1) 0.2 μg ml^{-1} tetracycline (Sigma, T7660), (2) no additive or (3) 200 nM OHT and 0.2 μg ml^{-1} tetracycline, corresponding to the repressed (or 0 h), induced (or High Myc) and Low Myc conditions, respectively[18].

Primary B cells from wild type and homozygous c-*myc*$^{f/f}$ conditional knockout animals[35] were purified by negative selection with the B Cell Isolation Kit (Miltenyi Biotech). After exposure to a recombinant tat-Cre protein (50 μg ml^{-1} for 1 h in optimem + 1% serum)[36] in order to induce deletion of the c-*myc*$^{f/f}$ allele, splenocytes were stimulated with LPS (50 μg ml^{-1}; Sigma, L6237) in DMEM and IMDM (ratio 1:1), 10% serum, 1% glutamine, 1% NEAA, 1% penicillin/streptomycin and 25 μM β-mercaptoethanol[37]. Deletion efficiency was checked by qPCR on genomic DNA.

3T9 c-*myc*$^{f/f}$ and 3T9^{MycER} fibroblasts were grown in DMEM medium supplemented with 10% serum, penicillin/streptomycin and 2 mM L-Gln. The role of endogenous Myc in 3T9 c-*myc*$^{f/f}$ cells was studied by deleting the c-*myc*$^{f/f}$ allele through activation of a conditional CreER chimaera in quiescent cells before serum stimulation, as previously described[8]. Deletion efficiency was checked by qPCR on genomic DNA. To generate the 3T9^{MycER} line, the 3T9 c-*myc*$^{f/f}$ cells were infected with a pBabe-Bleo retrovirus[38] encoding the MycER chimaera[39]. Cells were selected with 400 μg ml^{-1} zeocin (Invitrogen, 46-0509) for one week and kept in zeocin-free medium for subsequent experiments. Single-cell clones were derived and tested for 'leakiness', that is, activation of MycER in the absence of OHT: a non-leaky clone (number 14) was used for all experiments described here. MycER was activated by the addition of 400 nM 4-hydroxytamoxifen (OHT) or ethanol (vehicle control). All cell lines were tested for Mycoplasma in our tissue culture facility before usage.

Chromatin immunoprecipitation. Fixation of cultured cells and their processing for chromatin immunoprecipitation (ChIP) were performed as described[40], except for blocking protein A-Sepharose beads with tRNA (Sigma) instead of salmon sperm DNA and purifying immunoprecipitated DNA through Qiaquick columns (Qiagen) instead of phenol–chloroform extraction. The following adjustments were made for *in vivo* analysis. MACS-sorted splenic B cells or dissected lymphomas were resuspended in PBS with 0.5% BSA and 2 mM EDTA at room temperature and fixed by addition of 1% formaldehyde for 10 min. Fixation was stopped by addition of 0.125 M glycine. Cells were washed three times in PBS, resuspended in SDS buffer (50 mM Tris at pH 8.1, 0.5% SDS, 100 mM NaCl, 5 mM EDTA and protease inhibitors) and stored at −80 °C before further processing for ChIP[40]. To minimize inter-individual variation we decided to process ChIP samples of control and pre-tumoral mice in pools of at least ten animals, however, due to the clonal nature of lymphomas we kept tumour samples separated. For ChIP-Seq analysis of Myc and RNAPII, lysates from 30 × 10^6 B cells or 50 × 10^6 3T9 fibroblasts were immunoprecipitated with 10 μg of the corresponding antibody (see below). For histone marks, lysates from 5 × 10^6 B cells or 3T9^{MycER} fibroblasts were immunoprecipitated with 3 to 5 μg of the corresponding antibody pre-bound to G-protein coupled paramagnetic beads (Dynabeads) in PBS, 0.5% BSA. After overnight incubation, beads were washed 6 times in a modified RIPA buffer (50 mM Hepes pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-deoxycholate) and once in TE containing 50 mM NaCl. DNA was eluted in TE/2% SDS and crosslink reversed by incubation overnight at 65 °C. DNA was then purified by Qiaquick columns (Qiagen) and quantified using PicoGreen (Invitrogen). Then 2–10 ng ChIP DNA was prepared for Solexa Genome Analyzer or HiSeq2000 sequencing with TruSeq ChIP Sample Prep Kit

(Illumina) following the manufacturer's instructions, except for the H3K4me3 ChIP-seq in P493-6 cells for which libraries were produced as published[41].

Antibodies. The following antibodies were used for ChIP: H3K4me1 (Abcam, ab8895), H3K27ac (Abcam, ab4729), H3K4me3 (Active Motif, #39159), Myc N262 (Santa Cruz, sc-764) and RNAPII N20 (Santa Cruz, sc-899). Normal rabbit IgG (Santa Cruz, sc-2027) was used as background control. All antibodies were ChIP-grade, as specified by the manufacturer. For western blot: Myc Y69 (Abcam, ab32072), Vinculin (Sigma, V9264).

RNA extraction and analysis. Total RNA was purified onto RNeasy columns (Qiagen) and treated on-column with DNase (Qiagen). Complementary DNA (cDNA) was produced using the reverse-transcriptase ImPromII (Promega). A total of 10 ng of cDNA was used for real-time PCR reactions with FAST SYBR Green Master Mix (Applied Biosystems). A total of 100 ng of total RNA were processed for NanoString analysis as described by the manufacturer. For RNA-seq, total RNA from 10^7 cells was purified using TRIzol (Invitrogen), treated with Turbo DNase (Ambion) and purified with RNA Clean XP (Agencourt). Then 5 μg of purified RNA was treated with Ribozero rRNA removal kit (Epicentre) and ethanol precipitated. RNA quality and removal of rRNA were checked with the Agilent 2100 Bioanalyser (Agilent Technologies). Libraries for RNA-seq were then prepared with the TruSeq RNA Sample Prep Kits v2 (Illumina) following manufacturer instruction (except for skipping the first step of mRNA purification with poly-T oligo-attached magnetic beads).

4sU labelling. Treatment of 3T9^{MycER} fibroblasts with 4-thiouridine (4sU, Sigma T4509), isolation and sequencing of 4sU-labelled RNA were performed as previously described[42] with minor modifications. Briefly, cells were labelled with 300 μM 4sU for 10 min. Medium was removed and cells were put on ice, washed 3 times, scraped and pelleted in ice-cold PBS. RNA was extracted with the Qiagen miRNeasy kit according to the manufacturer's instructions including the recommended DNase I digest. Total of 50 μg of total RNA was used for the biotinylation reaction. RNA was diluted in 100 μl of RNase-free water. 100 μl of biotinylation buffer (2.5× stock: 25 mM Tris pH 7.4, 2.5 mM EDTA) and 50 μl of EZ-link biotin-HPDP (1 mg ml^{-1} in DMF; Pierce/Thermo Scientific 21341) were added and incubated for 2 h at room temperature. RNA was precipitated and unbound biotin-HPDP was removed by a combination of chloroform/isoamylalcohol (24:1) precipitation with purification using MaXtract high density tubes from Qiagen. Biotinylated RNA was purified using Dynabeads MyOne Streptavidin T1 (Invitrogen). Before addition of RNA, 50 μl of beads were washed 2 times in washing buffer A (100 mM NaOH, 50 mM NaCl) and once in washing buffer B (100 mM NaCl). Beads were resuspended in 100 μl of buffer C (2 M NaCl, 10 mM Tris pH 7.5, 1 mM EDTA, 0.1% Tween-20) to a final concentration of 5 μg μl^{-1}. RNA was added in an equal volume and rotated at room temperature for 15 min. Beads were washed 3 times with washing buffer D (1 M NaCl, 5 mM Tris pH 7.5, 0.5 mM EDTA, 0.05% Tween-20). RNA was eluted from the beads in 100 μl of 10 mM EDTA in 95% formamide (65 °C, 10 min). RNA was extracted with the RNeasy MinElute Spin columns from Qiagen according to the manufacturer and eluted in 14 μl of RNase-free water. RNA quality was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies). The 4sU-sequencing library was prepared with the TruSeq RNA Sample Prep Kits v2 (Illumina) following the manufacturer's instructions starting from the RNA fragmentation step.

DNase I hypersensitivity. Genome-wide sequencing of DNase I hypersensitive sites (DNase I seq) was performed as described[43,44]. Briefly, 3T9^{MycER} fibroblasts were treated with OHT for 4 h. Cells were washed with PBS, trypsinized, pelleted (1,300 r.p.m., 5 min, 4 °C) and washed once more with PBS. Pipetting in the following steps was performed with cut tips to avoid DNA breaks due to pipetting force. Cells were resuspended in buffer A (15 mM Tris-HCl pH 8, 15 mM NaCl, 60 mM KCl, 1 mM EDTA pH 8, 0.5 mM EGTA pH 8, freshly supplemented with 0.5 mM spermidine and 0.15 mM spermine). An equal volume of lysis buffer (buffer A with 0.1% NP-40) was added and the cells were incubated on ice for 10 min. Nuclei were pelleted, washed once with buffer A and then resuspended at a concentration of 50 × 10^6 nuclei per ml. Then 10^7 nuclei were diluted with an equal volume of 2× DNase I reaction buffer (Roche). DNase I (Roche, 04716728001) was added at increasing concentrations (0, 100, 200, 300, 400, 500 U ml^{-1}) and DNA was digested for 10 min at 37 °C. An equal volume of Stop buffer (50 mM Tris-HCl pH 8, 100 mM NaCl, 0.1% SDS, 100 mM EDTA pH 8, freshly supplemented with 0.5 mM spermidine, 0.15 mM spermine and 10 μg ml^{-1} of RNase A) was added. Samples were incubated at 55 °C for 30 min (220 r.p.m. agitation). Then 0.2 μg μl^{-1} of proteinase K was added and samples were incubated at 55 °C overnight (220 r.p.m.). DNA was extracted using a standard phenol–chloroform extraction protocol, dissolved in 100 μl of TE (55 °C, 2 h). Then 300 ng of DNA of each digested sample was checked on an agarose gel for the appearance of a smear of slightly digested DNA. Small molecular weight DNA was purified using AMPure beads (Agencourt AMPure XP Reagent, A63881). The digested DNA samples (100 μl) were supplemented with 50 μl of AMPure beads, 150 μl of 20% PEG buffer (20% PEG8000, 2.5 M NaCl) and incubated for 15 min at room temperature. Beads were separated

on a magnet, washed twice with 80% ethanol and small molecular weight DNA was eluted in 100 μl of 5.5% PEG buffer. The eluted DNA was purified once more (20 μl of beads; 120 μl of 20% PEG buffer) and after washing eluted in 20 μl of H₂O. DNase I performance was checked by qPCR and samples for sequencing were selected based on the highest signal-to-noise ratio based on selected genomic regions (with 200–300 U ml⁻¹ of DNase I). Chosen samples were size-selected on an agarose gel, small molecular weight DNA (<500 bp) was eluted from the gel with a Qiagen Gel purification kit according to the manufacturer's instructions. Up to 10 ng DNA was prepared for HiSeq2000 sequencing with TruSeq ChIP Sample Prep Kit (Illumina) following the manufacturer's instructions.

Primer design and list of primers. Primers for ChIP and mRNA analysis were designed by using computer assisted primer design software (Primer3). The complete list of primers used in this study is shown in Supplementary Table 3.

Immunoblot analysis. 5×10^6 to 10×10^6 B cells were lysed with RIPA buffer (20 mM HEPES at pH 7.5, 300 mM NaCl, 5 mM EDTA, 10% glycerol, 1% Triton X-100, supplemented with protease inhibitors (Mini, Roche) and phosphatase inhibitors 0.4 mM ortovanadate, 10 mM NaF) and sonicated. Cleared lysates were electrophoresed and immunoblotted with the indicated primary antibodies. Chemiluminescent detection, after incubation of the membranes with appropriate secondary antibodies, was done through a CCD camera using the ChemiDoc System (Bio-Rad). Quantification of protein levels was done using the Image Lab software (Bio-Rad, version 4.0).

Proliferation and cell size analysis. To measure cell size, 500,000 live cells were resuspended in 500 μl of PBS and 40,000 total events were collected using a FACSCalibur machine (Becton Dickinson). Propidium iodide (PI) staining solution was added to exclude dead cells from the analysis. Data were then analysed by using FlowJo software (TreeStar) and the mean of PI negative population scored. BrdU incorporation was analysed as described[45]. Cell proliferation was monitored using the CellTiter-Glo Luminescent Cell Viability Assay (Promega).

NanoString analysis. For quantitative mRNA measurements on the NanoString platform[46], we used four nCounter Reporter CodeSets. (1) A custom CodeSet was used for monitoring gene expression in the Eμ-*myc* model (Supplementary Table 4, Fig. 2f), for which we selected 754 genes (among which 458 were bound at their promoter by Myc in T) covering the whole expression range and regulatory patterns seen by RNA-seq, including 25 genes classified as non-expressed and 5 housekeeping genes (*Crocc, Sdha, Tbp, Tubb1* and *Tubb4*). (2) A custom CodeSet was used for monitoring gene expression in 3T9 c-*myc*ᶠ/ᶠ cells (Supplementary Table 5, Fig. 3f), with 446 genes covering the whole expression range mapped by RNA-seq in 3T9^MycER fibroblasts, and including 30 Myc-dependent serum response (MDSR) genes and 20 Myc-independent serum response (MISR) genes[8]. (3) A custom CodeSet was used for monitoring MycER-responsive genes in 3T9^MycER cells (Fig. 3d). This CodeSet includes the following 55 genes: *Arntl, Ddx58, Olfml2b, Ypel5, Lasp1, Vwa5a, Hsd17b11, Clec2d, Ctso, Prmt2, Myc, Capg, Crocc, Ubb, Dusp6, Rplp0, Car12, Tbp, Mycn, Cdca7l, Hpdl, Ifrd2, Hapln4, Efna3, Polr1b, Slc16a1, Slc19a1, Elovl4, Tfrc, Nnolc1, Wdr73, Zc3h8, Polr3g, Adi1, Fam136a, Bzw2, Wdr55, Taf4b, Mars2, Rrp9, Rragb, Slc25a33, Pdxp, Ica1, Smpdl3b, Dyrk3, Dgat2, GluI, Ifi30, Nr1d1, Reep6, Slc38a3, St6galnac4, Ankrd6* and *Smtnl2*. Data for these genes are provided in exactly the same order (sorted from left to right) in Fig. 3d. (4) A pre-designed NanoString CodeSet, the Human Cancer Reference Kit (GXA-CR1), was used for the experiments with P493-6 cells (Extended Data Fig. 6c).

Dedicated nCounter software was used for data analysis, and raw counts were normalized on the geometric mean of the internal positive control probes included in each CodeSet. Data were plotted either without further normalization, or normalized to cell equivalents (based on the total RNA recovered per cell in each single sample).

Statistical analysis. All the experiments were performed on biological replicates unless otherwise specified. Sample size was not predetermined and is reported in the respective figure legends. Two-tailed Student's *t*-test was used to calculate *P* values; significant values are specified in the figure legends.

Computational analysis

NGS data filtering and quality assessment. ChIP-seq and RNA-seq NGS reads sequenced with the Illumina HiSeq2000 were filtered using the fastq_quality_trimmer (setting the options to −Q33 −t 20 −l 10) and fastq_masker (setting the options to −q 20 −r N) tools of the FASTX-Toolkit suite (http://hannonlab.cshl.edu/fastx_toolkit/). Their quality was evaluated and confirmed using the FastQC application (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Analysis of ChIP-seq data. ChIP-seq NGS reads were aligned to the mm9 (Eμ-*myc* and 3T9 fibroblasts data) and hg19 (P493-6) genomes through the BWA aligner using default settings[47]. Peaks were called using the widely used MACS software (v1.4)[48]. Only peaks with *P* values $< 1 \times 10^{-8}$ were retained (positive peaks). MACS was also used to perform saturation analysis (as a control of false negatives) and to determine an estimated false discovery rate (as a control of false positives) for each experiment. In the saturation analysis, the fraction of peaks confirmed

with 80% of the reads was determined. For most of the samples this fraction was higher than 60%. False discovery rate was determined as the proportion of negative vs. positive peaks. Negative peaks were identified by calling MACS on the input samples, using the ChIP as reference. False discovery rates were typically lower than 5–10%.

Normalized reads count within a genomic region was determined as the number of reads per million of library reads (total number of aligned reads in the sequencing library). Peak enrichment was determined as $\log_2 (\text{ChIP}_w - \text{input}_w)$, in which ChIP_w and input_w is the normalized count of reads in the peak region in the ChIP and in the corresponding input sample. Myc P binding sites are defined as the union of peaks over 3 ChIP-seq experiments on independent P samples. Enhancers are defined as distal H3K4me1 peaks, that is, peaks not overlapping with promoters (−2 kb to +1 kb from TSS) and not associated with CpG islands (CGIs).

The RNA polymerase II stalling index (SI, also called elongation rate)[30,49] was calculated as SI = Prom/GB; prom refers to the read counts on the promoter (TSS ± 300 bp interval) and GB to the read counts in the gene body (the interval between TSS +301 and 3,000 bp after the TSS). These values were normalized both to library size (total number of reads) and to the length of the interval, and only genes with GB > 600 and with a RNAPII ChIP-seq peak in the region [TSS −2,000; TSS + 1,000 bp] were considered.

RNA-seq data analysis. RNA-seq NGS reads were aligned to the mm9 mouse reference genome using the TopHat aligner (version 2.0.6) with default parameters[50].

For Eμ-*myc* data read counts were associated to each gene (based on UCSC derived mm9 GTF gene annotations) using the HTSeq software (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html) setting the options -q --mode=intersection-nonempty --stranded=no. Absolute gene expression was defined determining RPKM as previously described[51]. Differentially expressed genes (DEGs) were identified using the Bioconductor[52] package DESeq[53] based on read counts, considering genes whose *q* value relative to the control is lower than 0.05 and whose maximum expression is higher than RPKM of 3.

For 3T9^MycER data, read counts were defined for each gene isoform as the sum of the read counts over all exons. Exon read counts were determined based on the TxDb.Mmusculus.UCSC.mm9.knownGene Bioconductor[52] annotation library using the countOverlaps Bioconductor[52] method. We calculated the read counts of a gene with more than one isoform as the rounded mean of the counts of its isoforms. The absolute expression of a gene in both RNA-seq and 4sU-seq was estimated through RPKM as previously described[51], defining total library size as the number of reads mapping to exons only. This is due to the fact that 4sU-seq library shows a marked enrichment in the reads mapping over intronic features compared to the RNA-seq library[42]. Therefore, in order to compare the expression of genes in both types of libraries, we measured library sizes using the reads mapping over exonic features only. DEGs were identified using the Bioconductor[52] package DESeq2, considering genes whose *q* value relative to the control is lower than 0.01 and whose minimum expression in all samples is higher than RPKM of 1.

For both Eμ-*myc* and 3T9^MycER RNA-seq data, samples were normalized using DESeq and DESeq2 based on scaling factors, assuming that most genes are not differentially expressed. For each gene the scaling factor is computed as the median of the ratio of its read counts over its geometric mean across all samples.

Other bioinformatic analysis. Bioinformatics and statistical analysis, including heatmaps of ChIP-seq data, GeneOntology enrichment, and hierarchical clustering of RNA-seq data were performed using R and Bioconductor packages. R scripts, data, and the R code used to reproduce bioinformatics and statistical analyses are available at (http://genomics.iit.it/supplementalData/SaboNature2014).

31. Adams, J. M. *et al.* The c-*myc* oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* **318**, 533–538 (1985).
32. Schmitt, C. A. *et al.* Dissecting p53 tumor suppressor functions *in vivo. Cancer Cell* **1**, 289–298 (2002).
33. Gorrini, C. *et al.* Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response. *Nature* **448**, 1063–1067 (2007).
34. Pajic, A. *et al.* Cell cycle activation by c-*myc* in a Burkitt lymphoma model cell line. *Int. J. Cancer* **87**, 787–793 (2000).
35. Trumpp, A. *et al.* c-Myc regulates mammalian body size by controlling cell number but not cell size. *Nature* **414**, 768–773 (2001).
36. Peitz, M., Pfannkuche, K., Rajewsky, K. & Edenhofer, F. Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. *Proc. Natl Acad. Sci. USA* **99**, 4489–4494 (2002).
37. Gerondakis, S., Grumont, R. J. & Banerjee, A. Regulating B-cell activation and survival in response to TLR signals. *Immunol. Cell Biol.* **85**, 471–475 (2007).
38. Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18**, 3587–3596 (1990).
39. Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G. & Evan, G. I. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res.* **23**, 1686–1690 (1995).

40. Frank, S. R., Schroeder, M., Fernandez, P., Taubert, S. & Amati, B. Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes Dev.* **15,** 2069–2082 (2001).

41. Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of *in vivo* protein–DNA interactions and epigenomic states. *Nature Protocols* **8,** 539–554 (2013).

42. Rabani, M. *et al.* Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature Biotechnol.* **29,** 436–442 (2011).

43. He, H. H. *et al.* Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. *Genome Res.* **22,** 1015–1025 (2012).

44. Sabo, P. J. *et al.* Genome-scale mapping of DNase I sensitivity *in vivo* using tiling DNA microarrays. *Nature Methods* **3,** 511–518 (2006).

45. Campaner, S. *et al.* Cdk2 suppresses cellular senescence induced by the c-*myc* oncogene. *Nature Cell Biol.* **12,** 54–59 (2010).

46. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnol.* **26,** 317–325 (2008).

47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

48. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

49. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nature Genet.* **39,** 1512–1516 (2007).

50. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).

51. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5,** 621–628 (2008).

52. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).

53. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).

54. Shaffer, A. L. *et al.* XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity* **21,** 81–93 (2004).

55. Ehrensberger, A. H., Kelly, G. P. & Svejstrup, J. Q. Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* **154,** 713–715 (2013).

56. Bouchard, C., Marquardt, J., Bras, A., Medema, R. H. & Eilers, M. Myc-induced proliferation and transformation require Akt-mediated phosphorylation of FoxO proteins. *EMBO J.* **23,** 2830–2840 (2004).

a

b

c

d

e

f

g

IgG
Expressed and Bound
Expressed and not Bound
not Expressed and not Bound

h

i

j
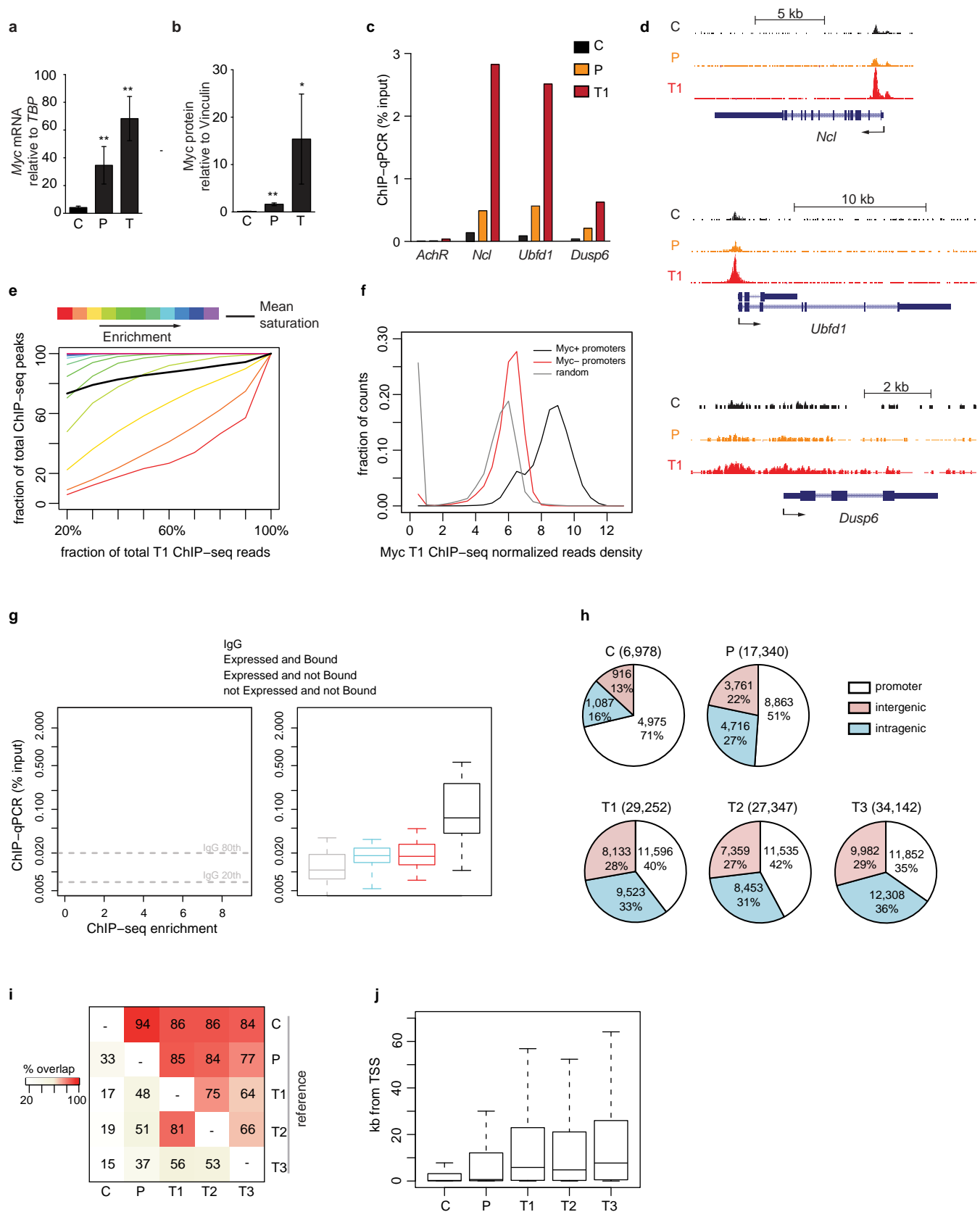
Extended Data Figure 1 | Intensity and frequency of Myc binding events increase during tumour progression, along with their distance from promoters. a, Quantification of c-*myc* mRNA levels by qRT–PCR (mean ± s.d. of $n = 6$, 8 and 6 for control (C), pre-tumoral (P) and tumoral (T) mice, respectively; Student's *t*-test). **b**, Myc protein levels were measured by quantitative immunoblotting (mean ± s.d. of $n = 4$, 4 and 3, for C, P and T mice, respectively; Student's *t*-test). **c**, qPCR analysis of Myc ChIP in C, P and T samples at the indicated promoters. *AchR* is a negative control and *Ncl*, *Ubfd1* and *Dusp6* are known Myc target genes. ChIP data are quantified as percentage of input DNA[40]. **d**, ChIP-seq tracks at the same target genes. The maxima of the tracks are 100 for *Ncl*, 50 for *Ubfd1* and 30 for *Dusp6*. **e–g**, ChIP-seq quality controls in the Eµ-*myc* data set. **e**, ChIP-seq saturation analysis for Myc peaks of different enrichment levels for a representative tumour (T1). In this analysis a random subset of the reads is removed (down-sampling) and the number of identified peaks in the down-sampled data set is compared to the full data set. If most of the peaks are confirmed upon down-sampling, the ChIP-seq is considered to be well saturated. **f**, Distribution of reads density (library size normalized) for Myc-bound or unbound promoters, as indicated (Myc+, Myc−). 20,000 random 3 kb regions not overlapping with promoters were used to assess background. The Myc signal in unbound promoters (red) is close to

the background (grey). **g**, Validation of ChIP-seq peaks by qPCR. We designed PCR amplicons at 42 sites that correspond to the following categories in Eµ-*myc*: (1) genes that are either not or barely expressed (RPKM < 4) and show no Myc binding in their promoter according to our ChIP-seq analysis (cyan), (2) genes that are expressed (RPKM > 4) but not bound by Myc (red), (3) genes that are both expressed and bound (black). An IgG antibody was used as negative control: the grey dotted lines indicate the 20th and 80th percentiles of the IgG signal. The left panel shows enrichment determined by ChIP-qPCR relative to that in ChIP-seq. ChIP DNA from C, P and T samples was used to amplify each test site. The right panel shows the distribution of signals obtained in ChIP-qPCR for the different categories of genes along with the signal obtained for all amplicons with the IgG antibody (grey). **h**, Pie charts showing the number and percentage of promoter, intergenic and intragenic Myc peaks in C, P and T samples (T1, T2, T3), based on the genomic position of each peak's midpoint. The number above each chart gives the total amount of peaks. **i**, Overlap between Myc ChIP-seq peaks. For each column the percentage of peaks overlapping (at least 1 bp) with the reference samples is reported. **j**, Box plot showing the net distance of Myc from the nearest TSS (upstream and downstream sites together). *$P < 0.05$, **$P < 0.001$.
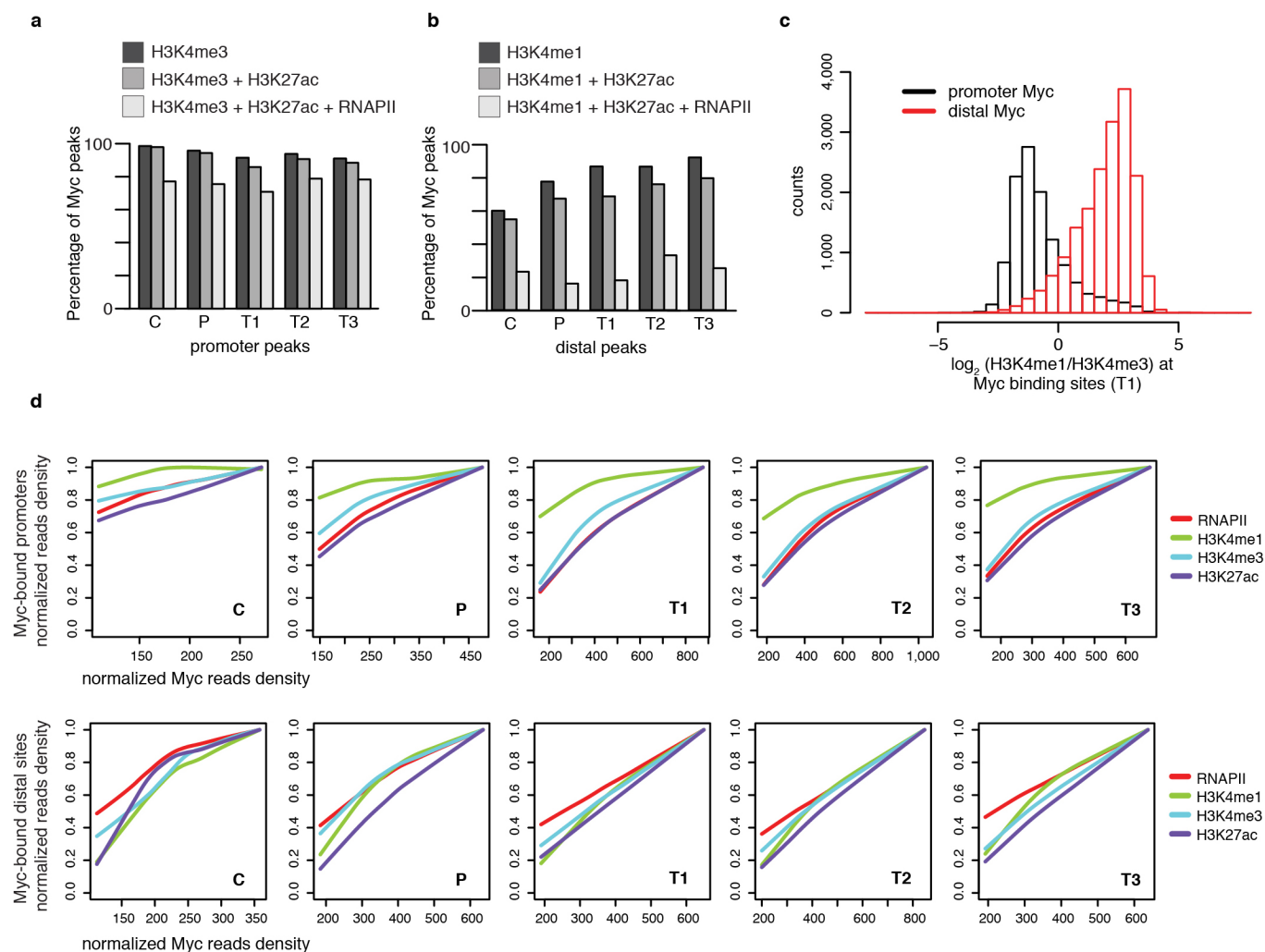
Extended Data Figure 2 | Myc binding to open chromatin. a, Percentages of Myc peaks associated with the indicated combinations of chromatin features at promoters. **b**, Same as panel **a** but for distal Myc binding sites. **c**, H3K4me1/ H3K4me3 ratios of library size normalized read counts over promoter-proximal (black) and distal (red) Myc-binding sites in a representative tumour (T1). **d**, Top, Lowess regression between the intensity of Myc binding (*x* axis) and level of RNAPII, H3K4me1, H3K4me3 and H3K27ac (*y* axis) at Myc bound promoters for C, P and T samples (T1, T2, T3); on both axes reads density is determined as the number of reads within 6 kb regions centred at the TSS. On the *y* axis data are divided by the maximum value observed for each mark. Bottom, same as above, for Myc-bound distal sites (6 kb centred at Myc peaks).
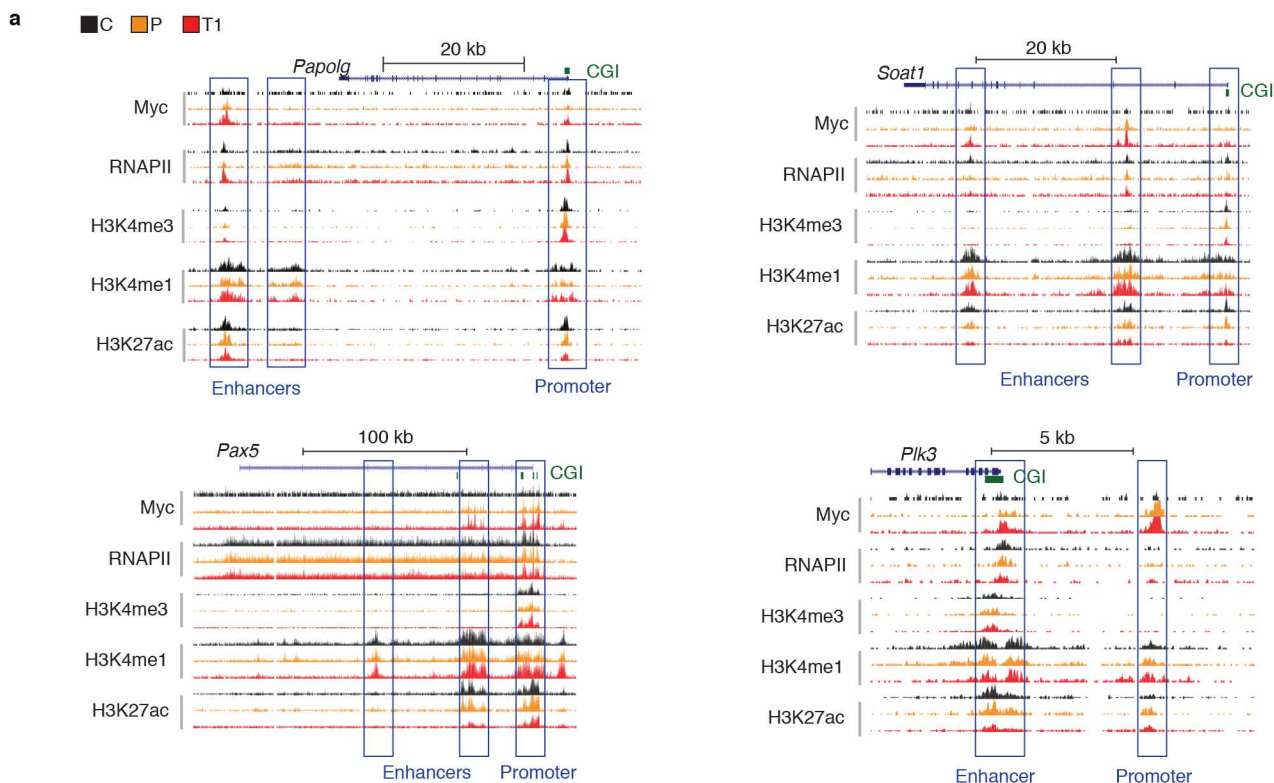
Extended Data Figure 3 | Myc binds active enhancers. a, Representative snapshots of Myc-bound regions in the genome. ChIP-seq tracks are shown for Myc, RNAPII, H3K4me3, H3K4me1 and H3K27ac in the indicated samples (C, P and T1 as a representative tumour). CGI, CpG island. Promoter and enhancer regions, on the basis of annotation and chromatin features, are indicated. **b**, List of the domains (amplicon) previously tested and validated (last column) as B-cell enhancers[14]. Bound C, P, T: Myc-binding status in our data set (y, yes; n, no). Note that the two domains listed in bold are also displayed next to the table (*Bst1* and *Bcl7a* loci).

a

b
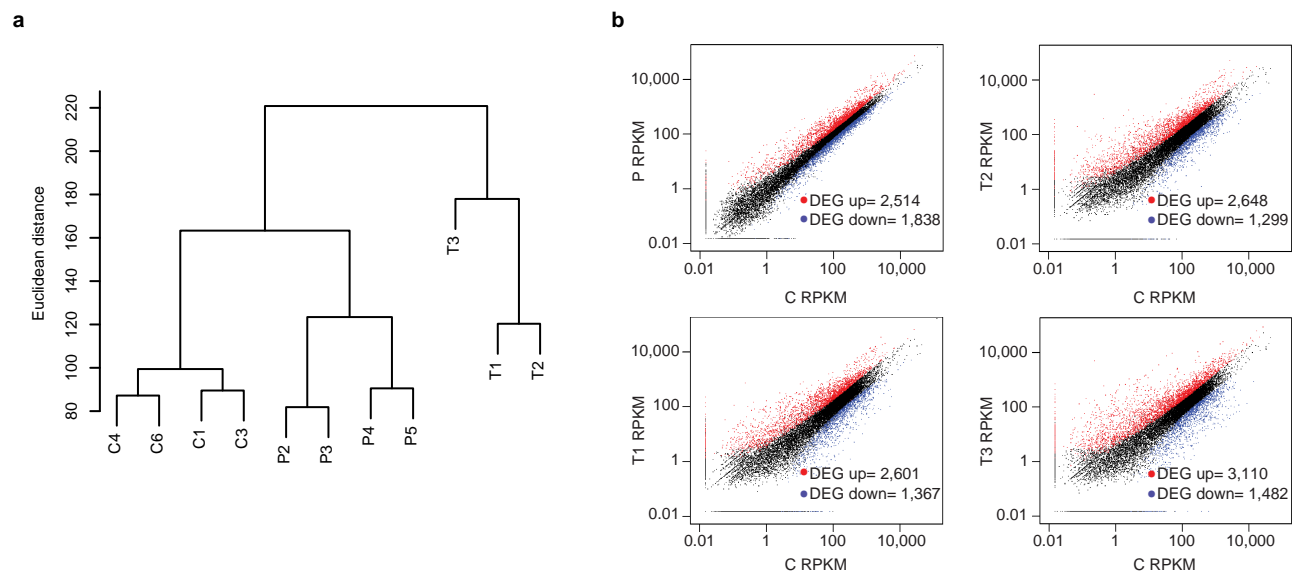
Extended Data Figure 4 | Gene expression data. a, Hierarchical clustering of C, P and T samples based on the RNA-seq data shown in Fig. 2a. C and P samples showed relatively homogeneous profiles, whereas T samples were more variable, consistent with the clonal nature of lymphomas. **b**, Scatter plots with differential gene expression; our calls for up- and downregulated DEGs are reported in red and blue, respectively, and their corresponding numbers reported as inset.
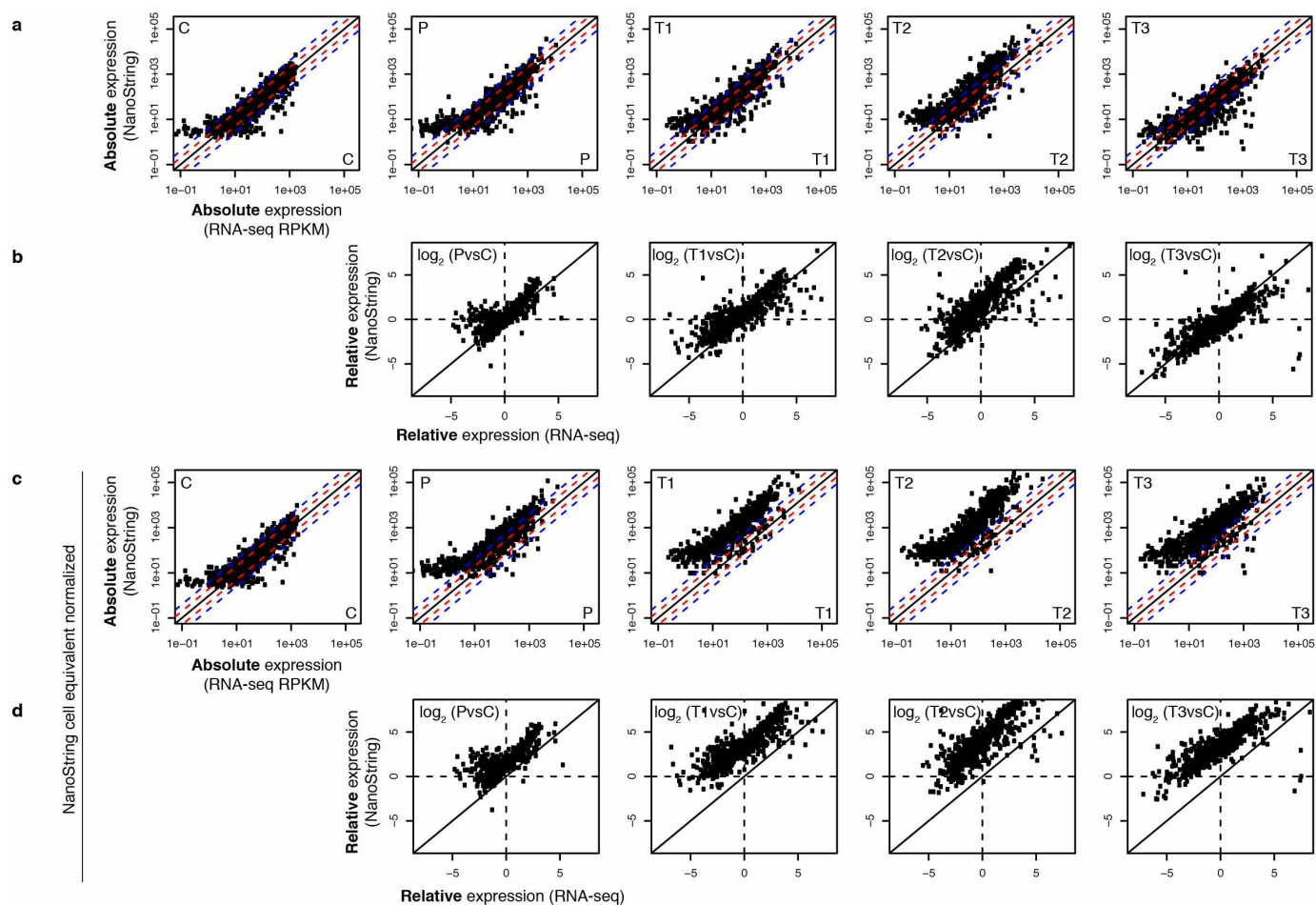
Extended Data Figure 5 | Comparison of NanoString and RNA-seq gene expression data. All scatter plots confront NanoString and RNA-seq data on the *y* and *x* axes, respectively. **a**, Absolute expression levels (NanoString counts versus RNA-seq RPKM units). Red and blue dashed lines indicate ± 2 and ± 4 fold differences, respectively. **b**, The data in panel **a** expressed as fold-change in the corresponding P and T samples relative to C ($\log_2$ FC: $\log_2$ of fold-change). **c**, Same as panel **a**, normalized to cell equivalents. **d**, Same as panel **b**, normalized to cell equivalents.
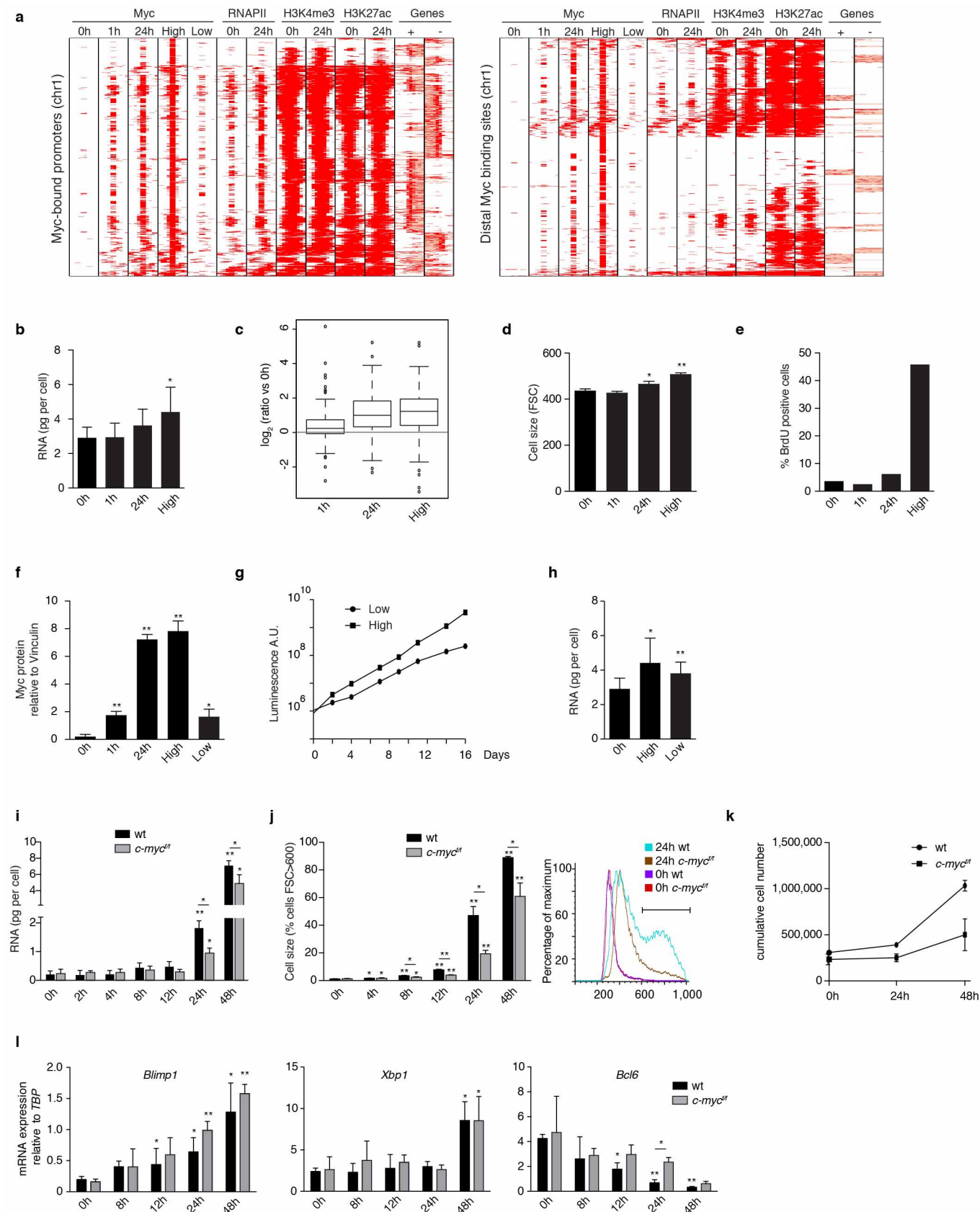
Extended Data Figure 6 | RNA amplification correlates with cell cycle entry and cell growth in P493-6 and primary murine B cells. a–h, P493-6 cells. Cells were cultured in tetracycline-containing medium (0 h: tet-Myc repressed) or Tet-free medium (tet-Myc on) for 1 h, 24 h or multiple passages (High Myc), as indicated. Low Myc: cells were passaged in Tet- and OHT-containing medium (see text). **a,** Chromatin profiles of Myc-bound promoters (left panel) and distal sites (right panel) on chromosome 1 (as in Fig. 1). **b,** Quantification of total RNA per cell (mean ± s.d., $n = 6$ per group; Student's t-test). **c,** Box plot showing gene expression changes relative to 0 h, as measured by NanoString normalized to cell equivalents ($n = 3$). **d,** FACS analysis of cell size distribution (mean ± s.d., $n = 3$ per group; Student's t-test). **e,** Percentage of BrdU positive cells (mean of two experiments). **f,** Measurement of Myc protein levels by quantitative western blot (mean ± s.d., $n = 3$ per group; Student's t-test). **g,** Growth curve (CellTiter-Glo assay; technical triplicate). **h,** Quantification of total RNA per cell (mean ± s.d., $n = 6, 6, 11$; Student's t-test). **i–l,** Primary murine B cells. Wild-type (wt) or c-$myc^{f/f}$ primary murine B cells were treated with tat-Cre (see Methods) and then with LPS for the indicated periods of time. **i,** Quantification of total RNA per cell (mean ± s.d., $n = 3$ pools of mice per genotype; Student's t-test). **j,** FACS analysis of cell size distribution (mean ± s.d., $n = 3$ mice per genotype; Student's t-test); the bar plot reports the percentage of cells with FSC higher than 600, as illustrated in the representative graph on the right. **k,** Growth curve, based on direct cell counts (mean ± s.d., $n = 3$ mice per genotype). **l,** qRT–PCR analysis of genes regulated during differentiation in plasma cells (mean ± s.d., $n = 3$ pools of mice per genotype; Student's t-test). *Blimp1* and *Xbp1* are induced, whereas *Bcl6* is repressed during plasma cell differentiation[20,54]. The data show that these regulatory patterns are intact in c-myc-deleted cells. P values are shown either relative to 0 h, or between wild type and c-$myc^{f/f}$ cells, as indicated. *$P < 0.05$, **$P < 0.001$.
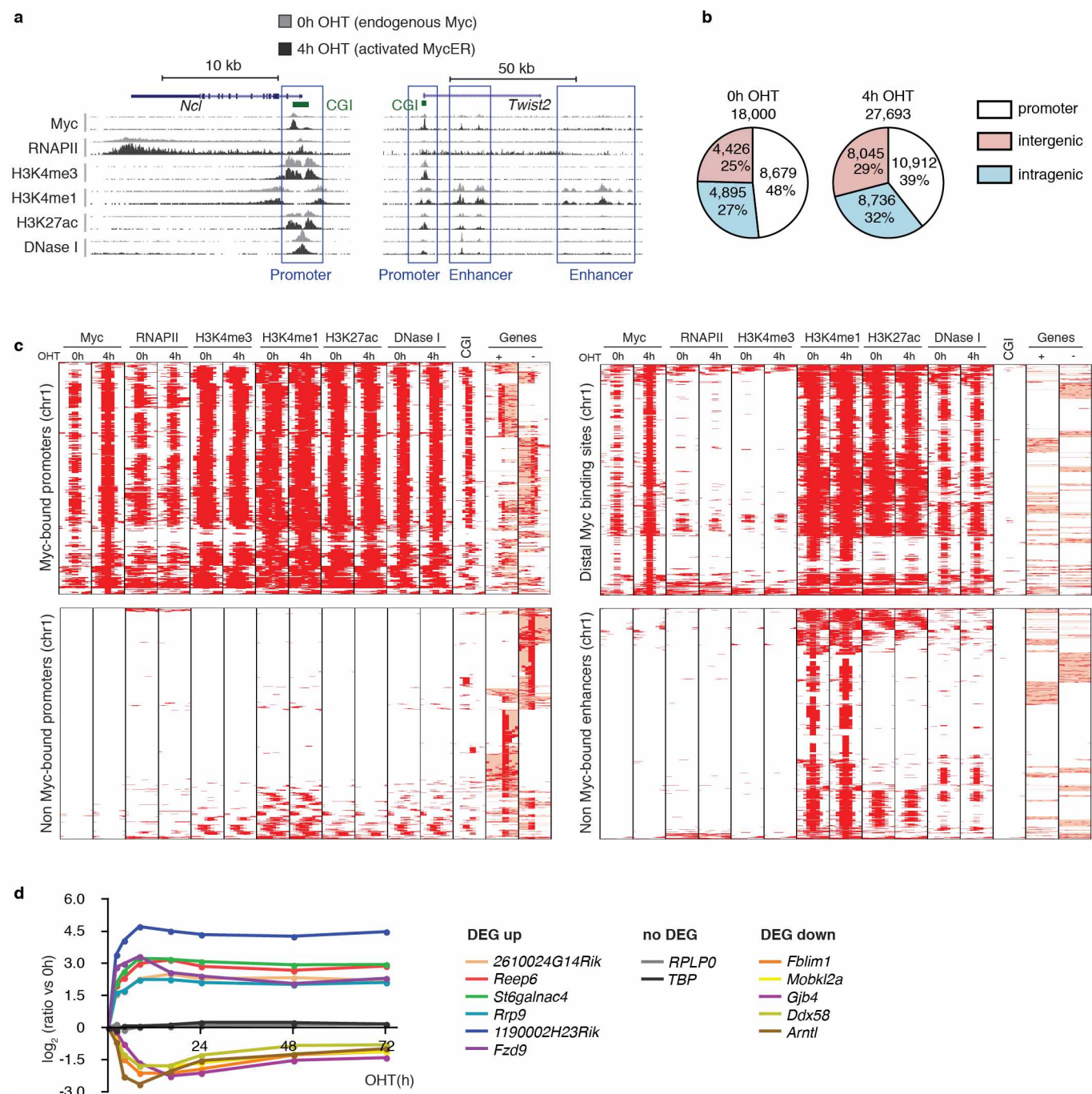
Extended Data Figure 7 | MycER binds active promoters and enhancers.
a, ChIP-seq tracks at two representative loci for the indicated proteins and histone marks as well as DNase I hypersensitive sites in 3T9^MycER fibroblasts before (0 h OHT, endogenous Myc) and after MycER activation (4 h OHT). Maxima in the Myc tracks are 80 for *Ncl*, and 20 for *Twist2*. CGI, CpG island. **b**, Annotation of Myc binding events, as in Extended Data Fig. 1h. **c**, Upper panels: chromatin profiles of Myc-bound promoters and distal sites on chromosome 1; lower panels: same at non-bound promoters and enhancers (as in Fig. 1). **d**, qRT–PCR validation on a per-cell-basis of up-, down- and non-regulated genes after OHT treatment in 3T9^MycER fibroblast relative to untreated cells (mean log_2 FC, *n* = 3). RNA was extracted from 50,000 cells at each time-point.
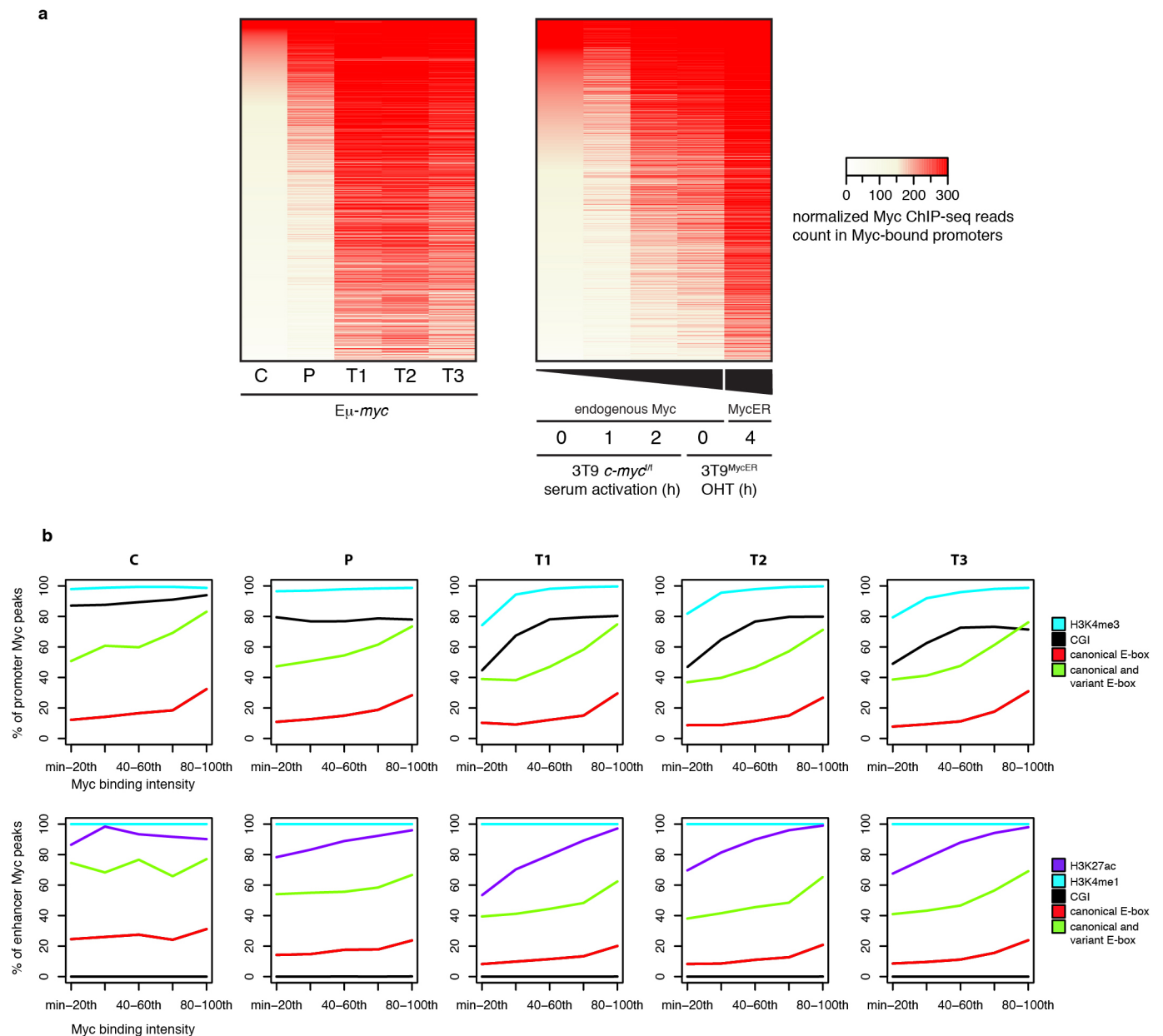
Extended Data Figure 8 | Myc binding hierarchy. a, Heatmaps showing library size-normalized ChIP-seq read counts for Myc-bound promoters in Eμ-*myc* and 3T9 samples, as indicated. Promoters are ranked from top to bottom by reads density. **b,** For all the Eμ-*myc* samples (C, P, T1, T2, T3) Myc binding events were divided in 5 equally sized groups based on their enrichment. Within each group the graphs report the percentage of Myc peaks overlapping (at least 1 bp) with the indicated histone marks, CpG islands (CGI), canonical and variant E-boxes or only the canonical E-box within a 100 bp window around Myc peak summits. The upper and lower graphs refer to Myc peaks on promoters and enhancers, respectively.
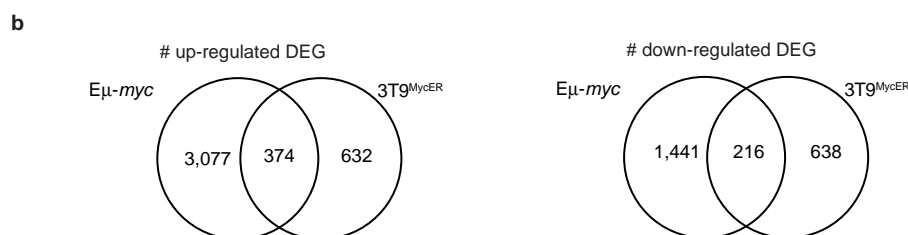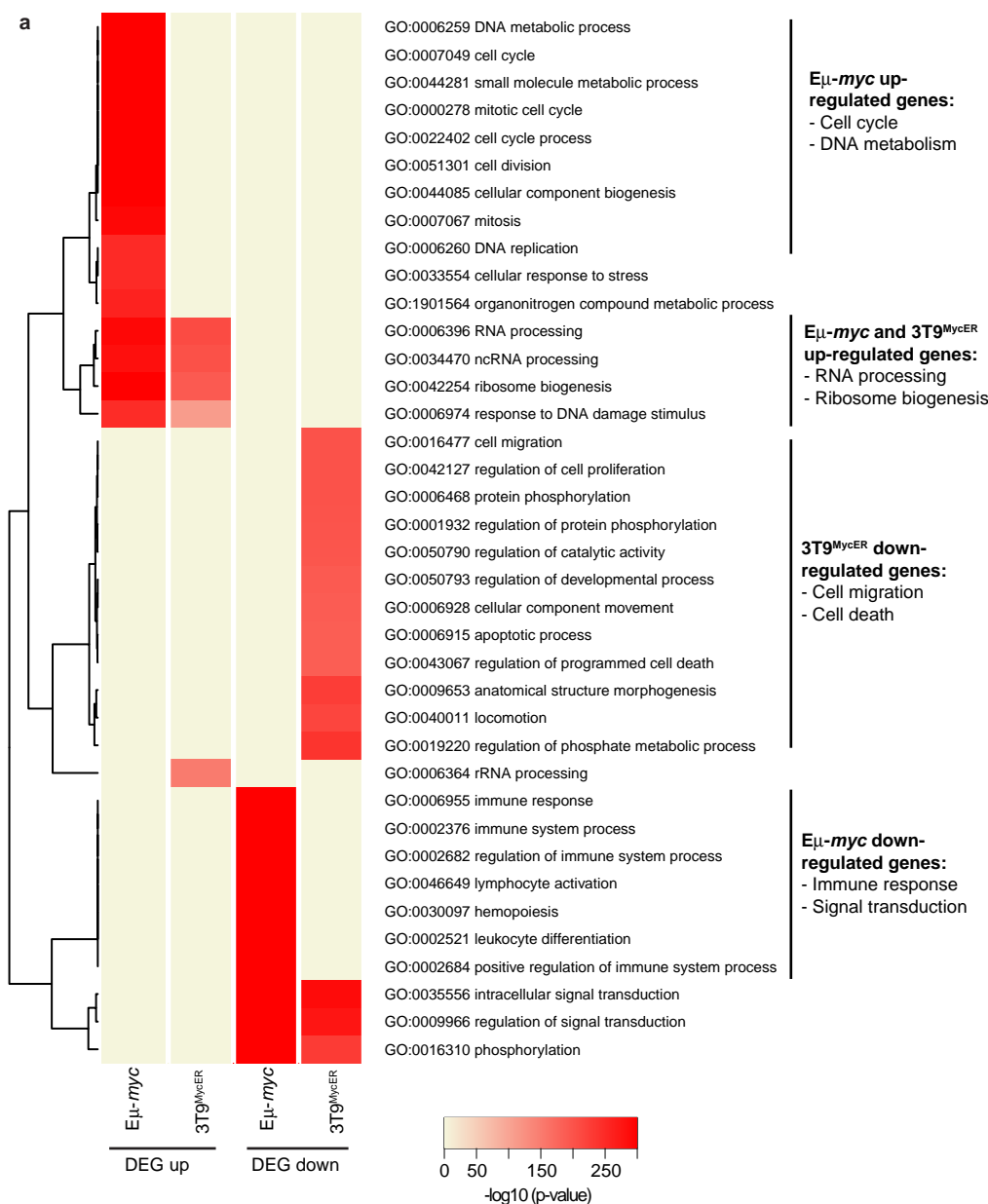
Extended Data Figure 9 | GeneOntology enrichment analysis.
a, GeneOntology (GO) terms enriched in up- and downregulated genes in Eμ-*myc* lymphomas (T) vs. C, and in OHT-treated (4 h) vs untreated 3T9^MycER cells. Genes differentially expressed in at least one tumour were considered for Eμ-*myc*. GO terms significantly enriched (Fisher test *P* value < 1 × 10^{-10}) in at least one gene set were selected (rows in the heatmap) and the *P* values for each

GO term in each gene set (columns) were colour-coded as indicated. Only GO terms in the biological process ontology that are assigned to less than 2,000 and more than 10 genes in the mouse genome were considered. **b**, Venn diagrams showing the overlaps among genes up- and downregulated in Eμ-*myc* lymphomas and 3T9^MycER fibroblasts.
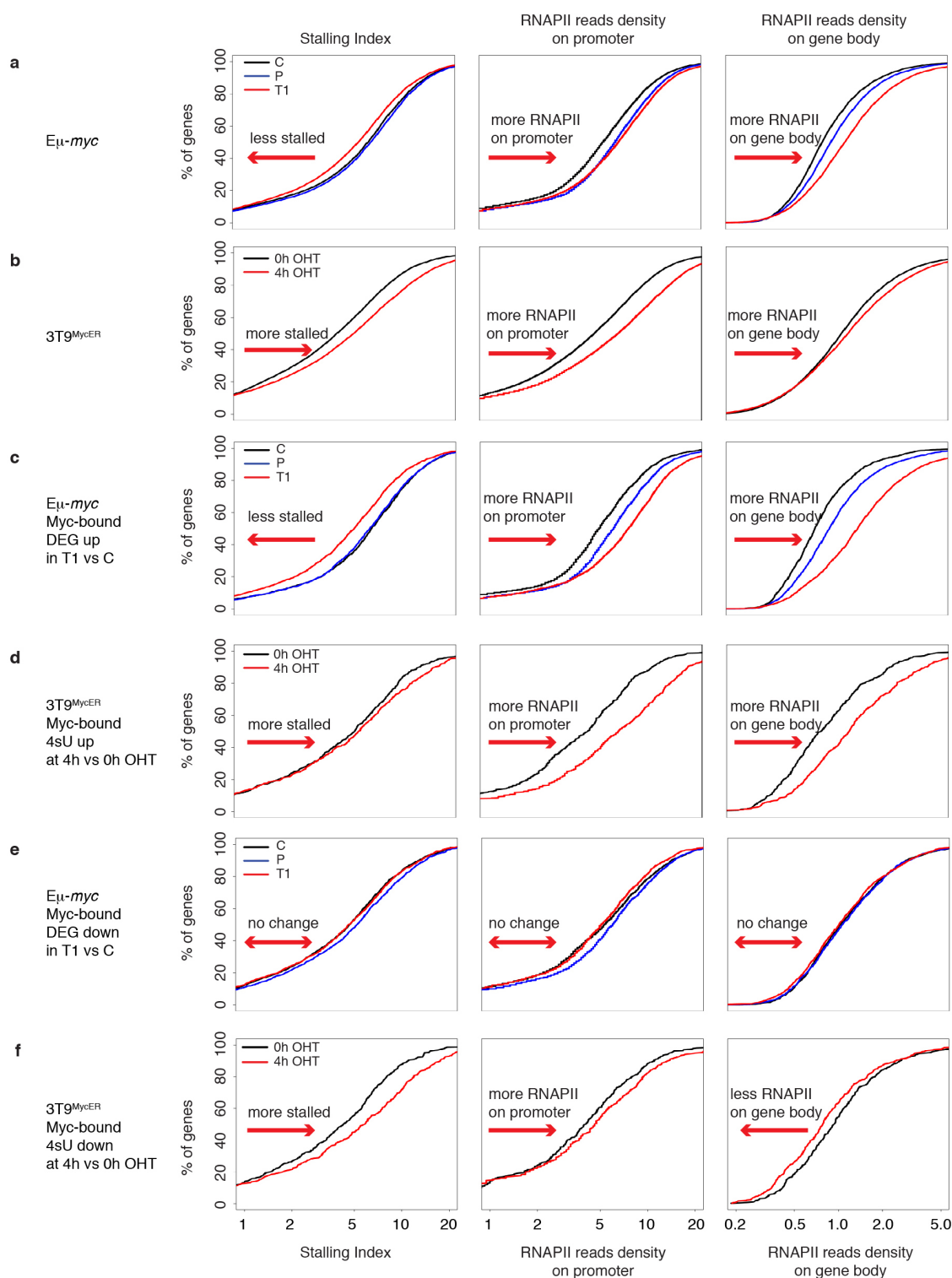
Extended Data Figure 10 | Stalling index analysis. The RNAPII stalling index is defined as the ratio of RNAPII read counts on promoter and gene body for all genes having an RNAPII peak on their promoters[30,49]. Each row shows the stalling index (left panels), the promoter (central panels) and the gene body (right panels) RNAPII counts for the indicated population of genes in the Eμ-*myc* and 3T9[MycER] models. **a**, **b**, All genes. **c**, **d**, Myc-bound upregulated genes (DEG up). **e**, **f**, Myc-bound downregulated genes (DEG down). For 3T9[MycER] cells, the 4sU-seq data were used, in order to consider only transcriptional changes, but similar results were obtained when using the RNA-seq data set (not shown). These observations can be summarized as follows. First, considering all genes, the stalling index followed different trends after Myc activation in different setting: a decrease along tumour progression in

Eμ-*myc* mice (**a**) and an increase upon MycER activation in fibroblasts (**b**). However, the density of RNAPII on the promoter and gene body revealed that both of these measures increased in both systems upon Myc activation, yet in different proportions, resulting in the opposite variations in stalling index. Thus, the stalling index, although capturing the ratio between RNAPII counts on promoter and gene body, conveys no information on the changes of either quantity per se, and may prove mechanistically misleading[55]. Second, on Myc-induced genes, the increases in RNAPII on promoter and gene body were both larger (**c**, **d**), suggesting that Myc can upregulate transcription by increasing both RNAPII recruitment (or re-loading) and elongation[30,56]. As expected, these effects were lost on Myc-repressed genes (**e**, **f**).

STOP THE MICROBIAL CHATTER

Bacteria can coat everything from thermal springs to teeth. Researchers are looking for antibiotics that can subvert the signalling that the microbes use to carve their niche.

D. G. DAVIES/BINGHAMTON UNIV.
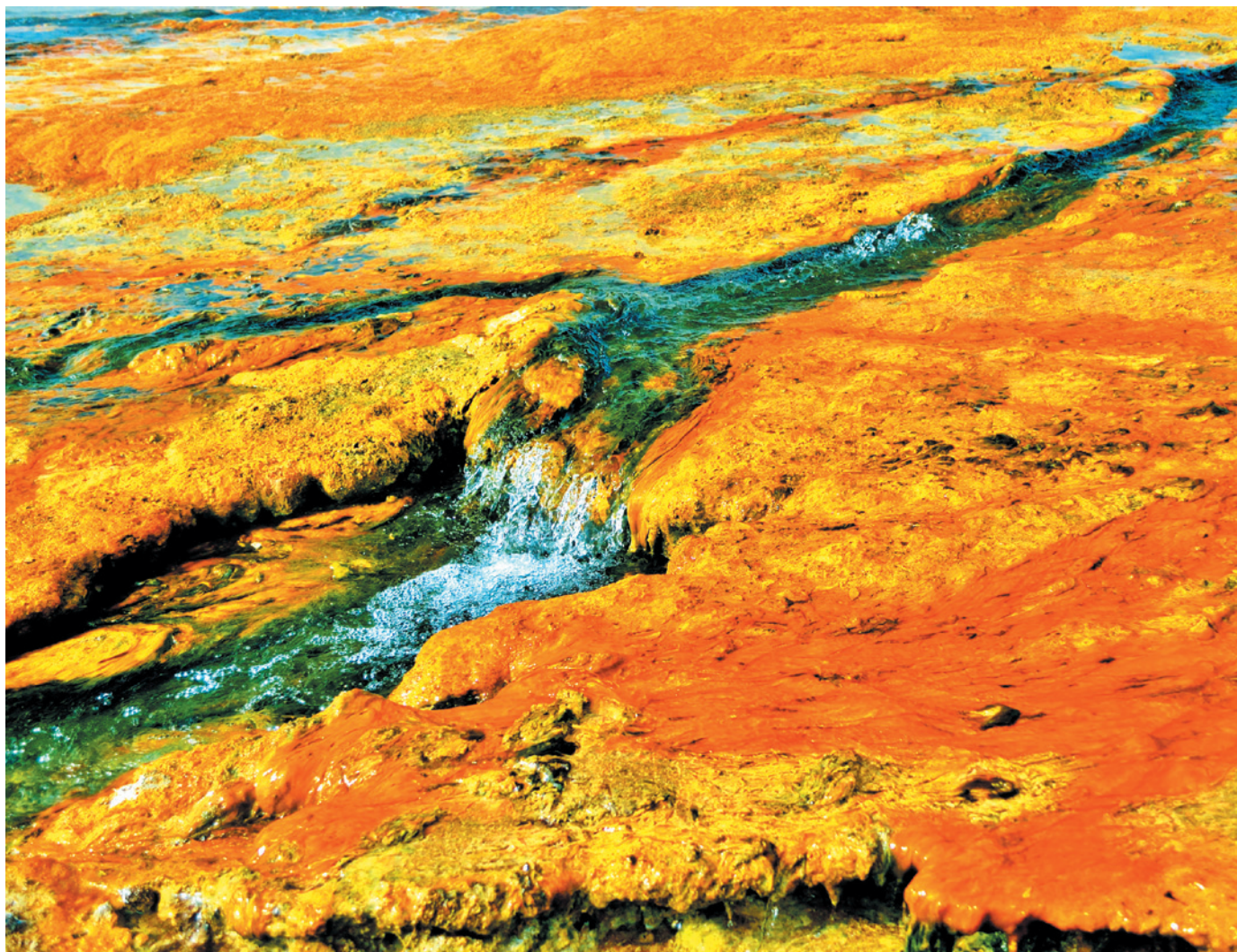
Sheets of communicating bacteria — or biofilms — are a common sight in the run-off channels from hot springs in Yellowstone National Park.

BY VIVIEN MARX

Bacteria are continually evolving ways to avoid the effects of antibiotics, and with the pipeline of new drugs drying up, infections are becoming more and more difficult to fight. As the need for innovative solutions grows, some microbiologists are teaming up with chemists and engineers to try to find ways to subvert the microbes by interfering with the signals they use to communicate.

To undermine the microbes' language, scientists first need to work out what they are saying. Bacteria use chemical signals to synchronize behaviour across a population. That behaviour can help us — in the digestion of food, say — but it can also kill us.

Such molecular coordination is thought to be central to the formation of biofilms — slimy mats of bacteria that spread across surfaces such as hospital catheters or water filtration systems. Some of the bacteria in a biofilm suspend their metabolism, explains microbiologist ▶

▶ Peter Greenberg of the University of Washington in Seattle, making antibiotics less effective because they tend to target bacteria that are still growing. The bacteria can also cover themselves in an armour made of polysaccharides and proteins that antibiotics find difficult to penetrate, says microbiologist Bonnie Bassler of Princeton University in New Jersey.

Such resistance to antibiotics can be treacherous, especially for people who have conditions such as cystic fibrosis that lead to long-term infections. Repeated treatments with broad-spectrum antibiotics heightens the risk that the bacteria will become resistant.

Bacterial communication was first studied in the 1960s, and not long afterwards, researchers found that a marine bacterium known as *Vibrio fischeri* would start to shine brightly once its population reached a certain density[1]. The finding that bacteria will turn their light on synchronously under certain conditions suddenly rendered bacterial behaviour visible and measurable, says Bassler. But because most scientists believed that bacteria were incapable of "fancy things" such as signalling, she says, the collective behaviour was generally dismissed as a "goofy phenomenon of bacteria living in the ocean".

Since then, researchers have observed this 'quorum-sensing' behaviour in many species[2–4] and have started to decipher the biochemistry and genetics of how it happens[5]. They have also been developing devices with which to characterize the messages that are transmitted and received.

In general, quorum sensing is triggered when signalling molecules emitted by individual bacteria pass a certain threshold, at which point the molecules bind to receptors on the bacteria and cause the entire population to express specific genes at the same time. In the case of pathogenic bacteria, the synchronized behaviour can include the release of molecules known as virulence factors, which help bacteria to colonize and harm their host. It also allows bacteria to create biofilms. As the organisms adhere to a surface, they keep signalling to one another. Once they sense a quorum, genes are upregulated and sticky exopolysaccharides are produced that 'glue' the bacteria together.

These findings initially led to excitement about the possibility of blocking infection by inhibiting bacterial communication. But the enthusiasm quickly waned when potential drugs failed in early-stage testing.

Now, scientists are taking a more sophisticated approach. The problem with the early work turned out to be in the assumption that the communication required only a few molecules, says Herman Sintim, a chemical biologist at the University of Maryland in College Park. The reality is much more complex, he says. "In human cultures, we all know that it does not take just one word to silence a crowd and so we should not expect that from our distant cousins, bacteria."

It has taken some time, but the research community in this field has grown and researchers have finally amassed enough knowledge about bacterial behaviour to start exploring how to stop the organisms from talking. "We are now getting there," says Bassler. Academics and companies are looking at fresh ways to study bacterial chatter and to create potential communication-disrupting drugs and agents for industrial and agricultural applications.

THE LANGUAGE OF BACTERIA

In developing drug candidates, researchers are sharpening their attack on infections beyond the broad-spectrum antibiotics currently in use. We need to talk to a specific bacterium "in a language only it understands", says Martin Blaser, director of the Human Microbiome Program at New York University Langone Medical Center. Narrow-spectrum antibiotics are less likely to engender resistance because they put fewer species under selection pressure. They also cause less disruption to the body's community of microbes — its microbiome. Broad-spectrum antibiotics will also remain necessary, especially for people who are very ill. In general, they are assumed not to have lingering effects, but Blaser says that "there's more and more evidence that's just not true". They could even wipe out microbial communities involved in the
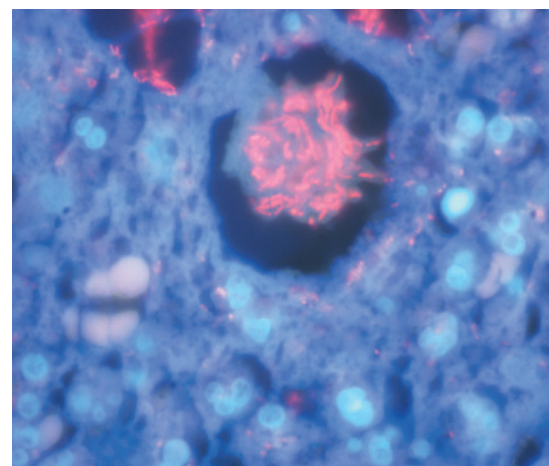
Infection-causing bacteria (red) are often buried deep in tissue and surrounded by white blood cells (blue), making them difficult to target.

developing metabolism of infants and children.

It might take some time, but research on bacterial communication will "without question" deliver therapeutic opportunities, says Ronald Farquhar, who directs research at Cubist Pharmaceuticals in Lexington, Massachusetts. Regulatory agencies are particularly open to drug-firm suggestions that will meet the needs of people with chronic infections, he says. For example, someone who needs to use a urinary catheter for a long period of time could take a low-dose agent to stop bacteria from forming a biofilm on the device.

Some drug candidates have already been identified. Microbiologists David Davies and Cláudia Marques from Binghamton University in New York, for example, have found a chemical that some bacteria make to address overcrowding[6]. The bacteria continuously produce *cis*-2-decenoic acid, a communication molecule. When the molecule reaches a critical threshold in a biofilm, a cascade of events is triggered, including changes in gene expression, prompting the bacteria to release themselves from the biofilm and disperse. Davies is now starting a company to commercialize a synthetic

---

SLUDGE FIGHT

Bacteria can be used to prevent biofilms from clogging the filtration membranes used in wastewater treatment.

1 As signalling molecules build up, the bacteria form a biofilm and clog the membrane.

Signalling molecule — Biofilm

2 Beads containing signal-digesting bacteria help to prevent biofilms.

Bead — Signal-quenching bacteria — Membrane

version of the acid for treating acne and disinfecting wounds.

But Greenberg, among others, thinks that caution is in order before moving potential therapies towards the clinic. Dispersing a biofilm could end an infection, he says, but it might also distribute it. "You might be making more trouble than you had to start with," he says.

Indeed, bacterial communication reveals ever more complexities. He has found, for example, that some bacteria in a community are cheats: they do not join the others in secreting enzymes in response to quorum-sensing signals, but still share in the benefits. "There are mixtures of cheats and cooperators in our laboratory experiments," Greenberg says, and a similar mix might be present in the infected lungs of a person with cystic fibrosis. Potential drugs could well be stymied by those cheats. Before developing therapies that disrupt communication, scientists need to know much more about quorum sensing and other bacterial behaviour, he says.

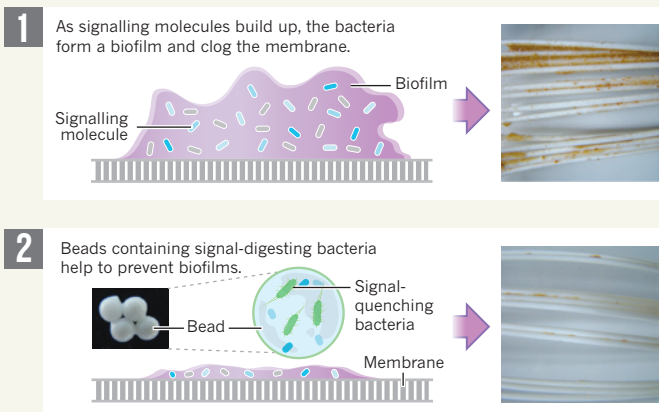Another complication is crosstalk between species and even across kingdoms. For example, Vanessa Sperandio, who studies bacterial communication at the University of Texas Southwestern in Dallas, has found that the stress hormones adrenaline and noradrenaline, which are present in the gut and elsewhere in the body, can amplify bacterial signalling and increase the virulence of *Escherichia coli* O157:H7 (ref. 7), a pathogen that causes bloody diarrhoea and can be fatal.

OTHER APPLICATIONS

To better understand the complexities of bacterial communication and how to use them against disease, the field is also turning to theoretical work, such as computational modelling and simulation, and to experiments with bacterial pathogens of plants. Greenberg and Lianhui Zhang, at the AStar Institute of Molecular and Cell Biology in Singapore, are working on a project funded by the Chinese government to use quorum-sensing inhibitors on crop pathogens. Such experiments could be proof-of-principle for biomedical applications, Greenberg says.

Quorum-sensing inhibitors could well make it to market in agriculture before biomedicine, says Paul Williams, a chemical biologist and pharmacologist at the University of Nottingham, UK, a hub for bacterial-communication research. Scientists and companies are also testing communication inhibitors for industrial applications. For example, microorganisms are being used in bioreactors to degrade the pollutants in wastewater. The water is then passed through a filter, but a build-up of bacteria can clog the pores of the membrane. The reactor then has to be taken offline, flushed out and cleaned with harsh chemicals such as chlorine — an energy-intensive process that

"There are mixtures of cheats and cooperators in our laboratory experiments."

The marine bacterium *Vibrio fischeri* glows brightly when it reaches a certain cell density, or quorum.

incurs more than half the cost of running a membrane bioreactor, says Chung-Hak Lee, a chemical engineer at Seoul National University.

Lee has come up with a potential solution. His approach taps into a typical communication network found in biofilms, in which enzymes secreted by some species digest signalling molecules emitted by others. He and his team isolated such signal-quenching bacteria and placed them in beads that contain pores that keep the bacteria in, but let signalling molecules pass through. When placed near the filtration membrane in a bioreactor, the beads undermine bacterial communication and help to stop biofilms from forming (see 'Sludge fight'). In lab tests and in a pilot-scale wastewater treatment plant, Lee has found that the beads save almost half of the energy costs of a conventional membrane bioreactor.

Several companies are exploring how to prevent biofilms for industrial and biomedical applications. Selenium, a spin-off company from Texas Tech University in Austin that is backed by the venture-capital firm Emergent Technologies, is developing selenium-containing coatings that could protect materials such as catheters, contact lenses and voice prostheses by producing reactive oxygen molecules that ward off bacteria.

Another company, Curza, founded last year in Salt Lake City, Utah, is developing coatings that prevent biofilms from forming on hip and knee implants. Its research involves chemical synthesis, molecular genetics, mass spectrometry and scanning electron microscopy, as well as a proprietary flow cell assay that better represents physiological conditions by using liquid flow rather than stagnant broth assays. The company says that the assay can help to

characterize whether a biofilm is prevented under real-life-like conditions and show what might happen as an antimicrobial compound dilutes away from a medical device's coating, for example.

And Kane Biotech of Winnipeg in Canada is developing combinations of antimicrobials and biofilm inhibitors for coating biomedical devices, treating wounds and protecting teeth and skin. Sri Madhyastha, chief scientific officer, says that one of their products has been licensed by a medical-device company. Kane also sells products through veterinarians and distributors, including a water additive aimed at preventing plaque from forming on the teeth of pets.

The company tried to obtain approval from the US Food and Drug Administration for an anti-biofilm enzyme in a wound-care product, but as a new chemical entity, it would require extensive testing. That route "is too expensive and time-consuming", Madhyastha says, so the company has put this product on the back-burner.

OBSERVATION PLATFORMS

To test their potential products, Kane's researchers use confocal microscopy and an instrument called the CDC Biofilm Reactor: a 1-litre beaker containing 8 slim rods around which liquid moves. Dotting the length of the rods are 24 circular disks on which biofilms can be grown and tested. The reactor was built under a licence from the US Centers for Disease Control and Prevention by BioSurface Technologies of Bozeman, Montana, which sells several other types of vessel in which scientists can grow and disrupt biofilms in a controlled, standardized environment.

At Fluxion Biosciences in South San Francisco, California, cell biologist Bryan Haines helps labs to set up the firm's BioFlux microfluidic platforms. The platforms allow scientists to do 24 biofilm experiments on one multiple-well plate. The temperature and gas content in the medium can be adjusted to suit the preferred growth conditions of the bacterium being studied. The wells are the reservoirs for reagents, potential antibiotics and bacteria; running underneath them are micrometre-scale channels in which a biofilm can grow. The plate is sealed at the top and users select the pressure with which to distribute fluids and cells through the channels, then observe the biofilm through an inverted microscope.

But Sintim says that scientists need better assays if they are to study the subtleties of bacterial communication. Cells live in a three-dimensional architecture and respond to many cues. And biofilms contain multiple species, making a specific biofilm hard to culture using traditional approaches. "Many systems that have been developed to date are reductionist systems," Sintim says, "and it is not obvious to me if data obtained from these reductionist platforms have any biological meaning."

Together with bioengineer William Bentley at his university, Sintim is developing a microfluidic system that will not just track cells moving through a three-dimensional space, but will also let experimenters perturb conditions and measure changes in appearance and behaviour. Their system uses a membrane to separate two types of bacteria. On one side of the membrane are bacteria they have engineered to fluoresce green under ultraviolet light. These bacteria secrete signalling molecules that can pass through the membrane. On the other side are bacteria engineered to fluoresce red only when they receive that signal. The device

Thomas Bjarnsholt wants assays that mimic the way that bacteria can be shielded from antibiotics.

allows researchers to alter the environment of each side independently and to control the rates of flow of liquids across the device (see 'Just watch'). Scientists can then study the effect of different gradients in a setting that is more typical of, for example, the body.

Thomas Bjarnsholt helps university-hospital physicians to diagnose infections and has a microbiology lab at the University of Copenhagen, where he is building a system for studying biofilms. Current assays do a poor job of showing how slowly a biofilm forms on a medical implant, he says, so he wants to develop an assay that more closely mimics the *in vivo* conditions. Also, only a few people develop infections when their hips or knees are replaced, so he hopes to determine what makes some luckier than others.

In his view, a communication disrupter should be tested not just by adding it to a

biofilm. In the chronically infected lung of a person with cystic fibrosis, antibiotics have to travel through the bloodstream, then diffuse through necrotic material, mucus and pus to get to the infection site. "It's all embedded in slime," he says. The slime also has anaerobic pockets, where antibiotics tend to fail. Just 40 micrometres of pus or mucus suffice to create such pockets. He is developing surfaces, gels and other media that mimic this kind of shielding and allow researchers to take this into account.

Quorum-sensing inhibitors and other communication disrupters will eventually emerge, Bjarnsholt predicts[8]. An area of interest for him is dressings, especially for people with diabetes, who repeatedly develop wounds. At the moment, dressings often contain silver, which acts as an antibacterial treatment, but infections still develop, so new approaches are needed, he says.

But new antibiotics will need more-expensive tests that require greater expertise to administer, says Sperandio. The standard way to test antibiotics is the minimal inhibitory concentration test, which measures the concentration at which a compound needs to be administered to stop bacteria from growing. Williams points out that this approach "is obviously of no use" for assessing compounds that disrupt communication.

In fact, says Sperandio, the whole communications approach to curing infection is at odds with the long-held dogma that a cure means killing the microbes. Communication disruptors could prevent pathogenesis without killing the pathogen, for example. Except in rare cases, such as infections of heart valves, it is not necessary to kill every bacterium, says Blaser. Even conventional antibiotics do not sterilize an organ; they reduce replication rates and "ultimately it is the immune response in patients that clears the infection", he says. New antibiotics could battle bacteria in this way, too.

The war on harmful bacteria is most definitely a war that humans need to win, says Blaser. But that does not mean we have to harm ourselves in the process. "We don't want a Pyrrhic victory," he says. ∎

Vivien Marx *is technology editor for* Nature *and* Nature Methods.

1. Nealson, K. H., Platt, T. & Hastings, J. W. *J. Bacteriol.* **104,** 313–322 (1970).
2. Engebrecht, J. & Silverman, M. *Proc. Natl Acad. Sci. USA* **81,** 4154–4158 (1984).
3. Hastings, J. W. & Greenberg, E. P. *J. Bacteriol.* **181,** 2667–2668 (1999).
4. Chen, X. *et al. Nature* **415,** 545–549 (2002).
5. Rutherford, S. T. & Bassler, B. L. *Cold Spring Harb. Perspect. Med.* http://dx.doi.org/10.1101/cshperspect.a012427 (2012).
6. Davies, D. G. & Marques, C. N. H. *J. Bacteriol.* **191,** 1393–1403 (2009).
7. Sperandio, V., Torres, A. G., Jarvis, B., Nataro, J. P. & Kaper, J. B. *Proc. Natl Acad. Sci. USA* **100,** 8951–8956 (2003).
8. Bjarnsholt, T., Ciofu, O., Molin, S., Givskov, M. & Høiby, N. *Nature Rev. Drug Discov.* **12,** 791–808 (2013).

JUST WATCH

Researchers at the University of Maryland are building a microfluidic device to study bacterial signalling. A membrane separates two types of bacteria — one fluoresces green and the other red — but allows the passage of signalling molecules. The device allows scientists to change the flow rate of liquids to look at concentration gradients, as well as to adjust various environmental factors, then observe how that affects communication between the bacteria.

Green-fluorescing bacteria emit signalling molecules.

Membrane

Waste

Bacteria fluoresce red when they receive signals from the 'green' bacteria.

Nutrients

Signalling molecules

CAREERS

GILLIAN BLEASE/IKON IMAGES/CORBIS

Equal access

Universities seek to recreate the success of one institution's mentorship programme for minorities in science.

BY VIRGINIA GEWIN

Astrophysicist Neil deGrasse Tyson, director of the Hayden Planetarium in New York and host of the US television show *Cosmos*, is arguably one of the best-known scientists in the United States — a success all the more notable given that he is an African American in a field dominated by white men. Tyson has suggested that the low numbers of minorities and women in the US

science workforce are due in large part to a lack of equal access to opportunities for entering that workforce.

To solve that problem, universities are now looking to the example of the Meyerhoff Scholars Program at the University of Maryland, Baltimore County (UMBC). It has crafted a formula for mentoring students from minority groups underrepresented in the sciences and helping to guide them into science, technology, engineering and maths (STEM) careers. And

that formula seems to work — African American Meyerhoff scholars are five times more likely than their counterparts at other US universities to pursue STEM PhDs (K. I. Maton *et al. Mt Sinai J. Med.* **79,** 610–623; 2012).

Now, with US$7.75 million in funding from the Howard Hughes Medical Institute (HHMI), Pennsylvania State University (Penn State) in University Park and the University of North Carolina at Chapel Hill (UNC) are trying to replicate the programme, which for two and a half decades has led the nation in the number of its minority graduates who go on to earn STEM PhDs.

Minorities account for almost 30% of the US population but make up just 9% of the nation's scientific workforce, according to the US National Institutes of Health (NIH). Only 1% of applications to the NIH for basic-research funding in 2000–08 were from African American researchers, and white applicants are nearly twice as likely as black applicants to win those awards (see *Nature* http://doi.org/b58f62; 2011).

Created in 1988, the Meyerhoff programme now has more than 900 alumni, of whom 184 have earned STEM PhDs and now work as university faculty members, government scientists or clinical physicians.

The Meyerhoff Adaptation Project at Penn State and UNC aims not only to implement the programme, but also to study how best to shape it for different campus cultures. "Other universities have tried to replicate individual components of the Meyerhoff programme, but none have been as successful (without that holistic approach)," says David Asai, senior director in science education at the HHMI.

The programme comprises several 'interventions' throughout a student's undergraduate career. It starts with a four-year financial-aid package and an intensive, six-week summer 'bridge' programme to prepare students for university study, and continues with study groups, tutoring, counselling and research internships throughout their undergraduate study.

The bridge programme is designed to help students to bond with each other and to prepare for the rigours of university courses. "For me, the summer bridge was pivotal," says Lekelia Jenkins, who was part of the fifth class of Meyerhoff students at UMBC and is now a marine-conservation scientist at the University of Washington in Seattle. She says that the friendships she formed in the programme helped to sustain her through the ups and ▶

downs of university and graduate studies. Meyerhoff scholars live together in residential housing during their first year at university, and are strongly encouraged to maintain study groups. The programme's staff advisers offer insights into coursework and research experiences; peers and graduates give moral support; and a network of Meyerhoff mentors provides career advice.

The programme also places students in summer research internships so that they can gain laboratory experience.

In a bid to improve the lacklustre numbers of minority undergraduates at Penn State and UNC who continue on for PhDs — fewer than ten a year at each — university administrators sought advice from UMBC biochemist Michael Summers, who helped to design the original Meyerhoff programme. Those discussions led to a joint proposal to the HHMI to study how the programme is implemented at both campuses, and the HHMI made the five-year funding commitment in May.

For many students, the adjustment to university is a challenge, which makes it all the more important to build an institutional culture to support minority scholars, says Mary Williams, associate dean for undergraduate education at Penn State. She says that the university is aiming to embrace the idea that excellence is accomplished through diversity, and that this project may help to eliminate stereotypes.

Meyerhoff graduate Lola Eniola-Adefeso, now a chemical engineer at the University of Michigan in Ann Arbor, says that UMBC's supportive culture may be difficult to replicate at other institutions. "UMBC started with the premise that minority students could achieve at an equal level to majority students, and all the faculty and staff fostered that sense of community," she says. She adds that the nurturing environment percolated down from the university's president through the rest of the campus. She scored top marks at UMBC, she says, "because I had that push to do well all around me".

Former Meyerhoff students often become research mentors themselves. Kafui Dzirasa, a neuroengineer at Duke University Medical Center in Durham, North Carolina, and a graduate of the programme, often brings Meyerhoff students into his lab as interns. Dzirasa stresses the importance of research experiences in helping students to decide

"The goal of the project is to learn from it so others are encouraged to try it."
David Asai

"For me, the summer bridge was pivotal."
Lekelia Jenkins

on a career path. After a successful internship at the University of Lancaster, UK, Dzirasa took a position at a petrol company. He realized that an industry post was not for him. "It was valuable for figuring out what I didn't want to do," he says. Now he enjoys helping current Meyerhoff scholars along on their academic journey.

There will be challenges in translating the programme to other institutions. Penn State and UNC will have to determine how to identify students who will succeed in the programme, says UMBC president Freeman Hrabowski. Applicants to UMBC's programme are nominated by secondary-school administrators, guidance counsellors or teachers. Finalists are selected on the basis of their intent to pursue an advanced STEM degree as well as their academic records, test scores, recommendation letters and community-service activities.

UMBC will work with UNC and Penn State to determine how to adapt the programme to fit each school's culture, demographics and research portfolio, and how to document how each aspect of the programme contributes to its overall success. Universities will have to assess, for example, the amount of academic preparation necessary for students to be successful. "If we're serious about getting other research universities to try programmes of their own, we have to look at each of the programme components to find out how and where they can work — and what adaptations need to be made," says Asai.

Once the UNC and Penn State programmes are fully implemented, each will endeavour to expand slowly to 35–40 students. Hrabowski and Asai hope that what they learn from the project will encourage other universities, foundations and agencies to adapt the programme.

"We in science education should be looking for dissemination of effective programmes — not every programme needs to be new," says Asai. "The goal is to learn from this so others are encouraged to try it."

Hrabowski agrees. "The vision is not simply creating a programme that encourages diversity in science but one that works to produce the best scientists in the world," he says. "That's what's revolutionary." ∎

Virginia Gewin is a freelance writer in Portland, Oregon.

Postdoc policies probed

Less than two-thirds of US institutions say that their postdoctoral researchers earn at least the US National Institutes of Health-recommended minimum of US$39,264, a survey has found. Health-care benefits are available to postdocs working as employees under a principal investigator's grant at some 95% of responding institutions, but less than two-thirds offer such benefits to those classified as trainees or on fellowships.

The US National Postdoctoral Association (NPA) in Washington DC, which represents about 70,000 postdocs across 190 US and Canadian institutions, this month released preliminary data from its Institutional Policy Survey. Building on a 2011 survey, the association polled 167 private, public and government institutions on postdoc policies concerning pay and benefits, career-development services and career-tracking practices, among others. It received full responses from 74 institutions.

No other agency or organization has produced similar information on postdocs. The NPA estimates that there are up to 91,000 postdocs in the United States, including those who are not NPA members.

The survey results showed that career-development services for postdocs vary widely. Most institutions (96%) offer training in writing grant applications, for example, but only just over three-quarters run programmes on presentation skills, and smaller proportions provide interview training, networking events or career-counselling appointments. Institutions with fewer than 750 postdocs offer especially scanty services. Some 57% of institutions say they have no postdoc handbook, and about one-fifth set no limit on how long a researcher can be classified as a postdoc. Just six institutions reported that they have a system to track postdocs' career arcs after they have left the institution.

Belinda Huang, executive director of the NPA, says that the association is preparing a report on the findings, for distribution to policy-makers, funders and university and institution administrations. It will call for extra training for careers outside academia, as well as more tracking of postdocs after they leave. "Now we have hard data," she says.

A LONG WAY FROM HOME

Distant memories.

BY SYLVIA SPRUCK WRIGLEY

Great-Aunt Gertrude chopped up the dried fruit while I stirred the batter. The fridge updated: we were 982,000 kilometres from Earth exactly. "Look, Mom, look," Lorissa shouted. She was waiting for all nines.

"Very exciting, pumpkin." I tried to look enthused.

Great-Aunt Gertrude was lecturing, as usual. "If you want to do it right, you must to mix it by hand, not using the electric mixers."

Lorissa crawled up on the stool, took the wooden spoon from me. "You are best helper," Great-Aunt Gertrude told her. "Much better than your mother. She's not careful when she stirs." Lorissa nodded, her chubby face serious as she carved careful paths into the batter. I held my tongue.

We didn't keep track of the Earth calendar on the ship. But Great-Aunt Gertrude had decided that this was a good time to celebrate *Weihnachten*. Never mind that no one had any presents or that there wasn't any point to wishing for peace on Earth and mercy mild. Lorissa decorated the lounge with cotton-ball snow and we sang all the carols and ate the lebkuchen. You didn't say no to Great-Aunt Gertrude.

Hugh sat in the comms room, as always. "Merry Christmas from Sittner735, is anybody out there?"

When Great-Uncle Herbert told us that he'd bought a fully laden escape shuttle, we all said he was crazy. He'd wasted his entire pension, fallen for scammers, gone senile. When the war started, he must have been pretty tempted to leave us all behind for laughing at him. But that would have left him alone for all eternity with Great-Aunt Gertrude, so he forgave us and made sure we all got on board. Thing is, no one knew how to drive the navigation system. Hugh, my husband, kept trying the radio in the hope of making contact with some other evacuees, but again, no one really knew what frequencies we should use or anything. So, there it was, six of us heading out into space, waiting to see how far this ship could go.

Every 28 cycles,

Great-Aunt Gertrude declared it was *Weihnachten* and the whole damn thing would start over. I knew it was every 28 cycles because she was right in sync with my period, so just as my back cramps kicked in, it was "come to the kitchen and help your baby stir the batter" like we hadn't just done this already.

This was the fourth Christmas celebration. Lorissa had just showed me the fridge display ticking over to 989,500 kilometres from Earth when Great-Aunt Gertrude collapsed. I carried her to her quarters, struck by how light she'd become in the few months since we left home. Great-Uncle Herbert ignored her, as always, sitting in the library studying his digital collection of star-system maps. Hugh called out into space, searching for signs of life. I sat with her every day. She had become small and frail overnight, her skin taut across her bones. I patted her hand and Lorissa brought her mugs of tea which she never drank. Sitting there day after day, I felt useless and angry and cold.

"Lorissa, go play with your father," I snapped after she asked me for the fifteenth time whether there was any lebkuchen left.

"He told me to come here," she whined. That was the final straw. I stormed to the radio room where Hugh sat, clutching the microphone as if it could save him. "Can't you just put that down for ten minutes and

take care of your daughter? Do I have to do everything?"

He turned and gave me a slow, hard glare. I stood my ground. "I'm trying to look after my aunt and Lorissa is constantly underfoot and all you do is sit there mumbling into the black like suddenly the Johnsons are going to turn up right next door. Don't you know it's hopeless?"

I wanted him to shout or rage or do *something* but he just stared at me and finally said in a low voice: "What do you want me to do?"

His cold face scared me more than anything else so far. "Help me with Great-Aunt Gertrude."

He shook his head. "How? She's out of pills, out of time."

"What?" My heart pounded as I realized the truth of it. "Who would be so stupid? Surely they must have bought spares, stocked up with a lifetime's supply when they bought the crates of spam and powdered ginger."

"Yeah, the doctor would totally write out the prescription for that," he said. He laughed. I stared at him, uncomfortably aware that I'd barely spoken to him in months. He was in here, day and night, calling out on the radio. "What's she going to do with all those pills, anyway?" He put down the mike and stepped close to me but didn't reach out. "Where do you think we are going, exactly?"

"Well, the ship's programmed, right? So we'll head out into space and then back again once the fall-out has cleared."

Hugh stared at me in that same way, as if he were looking at a stranger. Then he just shook his head and sat back down. "Sittner735, does anyone read?"

Lorissa ran into the comms room. "Is it *Weihnachten* again soon?" I took her to the kitchen and got the bowl down. "We'll make lebkuchen for Great-Aunt Gertrude," I told her. "Every day, for as long as we can."

The fridge ticked over: 997,799 kilometres. And only now had I realized that there was nowhere for our family to go. ∎

Sylvia Spruck Wrigley *was born in Germany and spent her childhood in Los Angeles. She now splits her time between South Wales and the Costa del Sol, two coastal regions with almost nothing in common. You can find out more about her at www.intrigue.co.uk.*

↻ **NATURE.COM**
Follow Futures:
🐦 @NatureFutures
f go.nature.com/mtoodm

natureOUTLOOK

ASSESSING SCIENCE

Lessons from Australia and New Zealand

natureOUTLOOK

ASSESSING SCIENCE

Cover art: Dale Edwin Murray

Editorial
Herb Brody, Michelle Grayson, Stephen Pincock, Kathryn Miller, Rebecca Dargie, Afsaneh Gray

Art & Design
Wesley Fernandes, Mohamed Ashour, Alisdair Macdonald, Andrea Duffy

Production
Karl Smart, Susan Gray, Ian Pope, Robert Sullivan, Chris Gilloch

Advertising
Kylie Ahern, Kate Aylett

Marketing
Hannah Phipps

Project Manager
Anastasia Panoutsou

Art Director
Kelly Buckheit Krause

Publisher
Richard Hughes

Magazine Editor
Rosie Mestel

Editor-in-Chief
Philip Campbell

Society benefits enormously from scientific research. We get new technologies, live longer and healthier lives, and gain deeper knowledge of our planet and the Universe. The issue of how to evaluate the fruits of research confronts scientists and policy-makers all over the world. Every country has its own set of circumstances surrounding its research infrastructure, wealth, and economic, environmental and developmental objectives — so there is no universal solution.

Earlier this year, at a symposium organized by *Nature* in Melbourne, Australia, a group of leading academics, funders and government advisers discussed how research outcomes are measured (see page S57). This Outlook supplement was influenced by these debates, although we at *Nature* take sole responsibility for its content.

As discussed at the symposium, both Australia and New Zealand have research assessment programmes that place heavy emphasis on research excellence (S52) — a qualitative determination that is heavily informed by quantitative metrics concerning, for instance, how often a paper is cited (S64). Both Australia (S67) and New Zealand (S82) have seen their global scientific standings rise in recent years — attributable at least in part to their assessment systems, even though Australia's system offers little financial reward (S81).

Measuring research using academic yardsticks largely ignores the wider impacts of research such as new policies or improved technologies. Academics and policymakers in both countries are considering the benefits and difficulties of trying to measure such impact (S72). Could the creation of 'citation equivalents' enable comparison of non-academic work against peer-reviewed literature (S77)?

We hope that the intense focus on these issues in Australia and New Zealand will inform and stimulate this crucial debate throughout the scientific world.

Philip Campbell, *Editor-in-Chief,* Nature
Michelle Grayson, *Senior Editor, Supplements*

CONTENTS

FUNDING BY NUMBERS

*Australia and New Zealand both rely on assessment schemes to improve research quality, yet the money associated with each is very different. By **Julie Gould**.*

The Excellence in Research for Australia defines research as:

"The creation of new knowledge and/or the use of existing knowledge in a new and creative way so as to generate new concepts, methodologies and understanding."

$1.3M
Northern Territory
2 universities

Queensland
8 universities
$14.2M

Western Australia
5 universities
$8.5M

Funding for universities as determined by ERA assessment[1].

South Australia
3 universities
$7.7M

New South Wales
10 universities
$28.6M

Australian Capital Territory
2 universities
$6.9M

Other (Australian Catholic University, across 4 states)
$0.3M

Victoria
9 universities
$24.1M

Tasmania
1 university
$2.3M

Darwin · Cairns · Townsville · Brisbane · Sydney · Canberra · Melbourne · Perth · Adelaide · Hobart

AUSTRALIA
RESEARCH ASSESSMENT FUNDING BY STATE

Australia's budget for higher-education research and development is an order of magnitude larger than that of New Zealand, yet the amount of money determined by Excellence in Research for Australia (ERA) is significantly smaller.

Population:
22M

MONEY MATTERS

The Excellence in Research for Australia (ERA) assessment comprises only 0.08% of all government funding for higher-education research and development (HERD)[1].

Overseas **$0.2B**
Business **$0.4B**
Donations **$0.2B**
Government (state, local, federal and competitive grants, inc. ERA) **$4.2B**
General University funds (e.g. student fees) **$5.1B**

$9.1B
Sources of funding

Experimental development **$0.7M**
Strategic basic research (new knowledge in specific areas) **$1.8B**
Pure basic research (new knowledge with no agenda) **$2.3B**
Applied research **$4.2B**

$9.1B
HERD expenditure

SCHOLARLY SPENDING

New Zealand continues to increase spending on HERD, but Australia is starting to pull back[2].

Percentage of GDP spent on HERD — Australia, New Zealand*; Year (2000–2010)

*Data for Australia are produced on even years; data for New Zealand are produced on odd years.
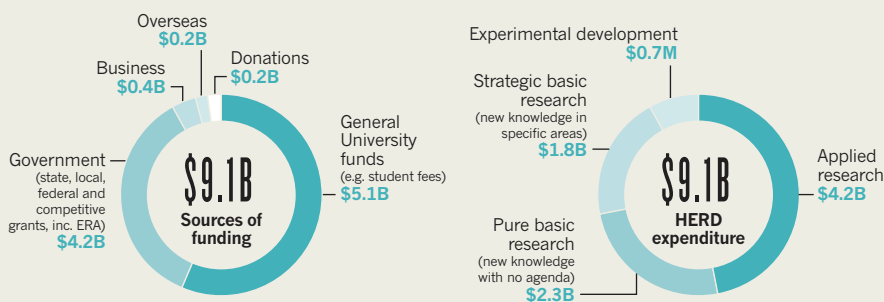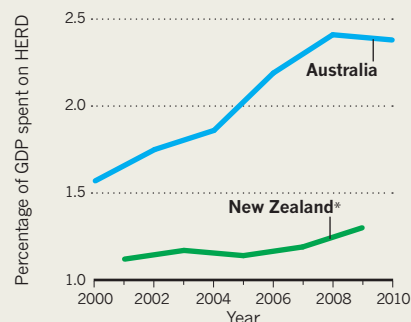
EXCELLENCE IN RESEARCH FOR AUSTRALIA: TIMELINE

Policy-makers and government officials have changed funding allocations from formula favouring quantity to assessment criteria favouring quality[3].

1990	1995	2000	2005	2010		2015
Australian Research Council established.	Research block funding became formula based, measuring number of papers published.		Minister for Education, Brendan Nelson, calls for new Research Quality Framework (RQF).	ERA pilot study. Excellence in Research for Australia (ERA) announced. RQF abandoned.	ERA 2012 report.	Next ERA evaluation.
Introduction of Unified National System, which was a framework to reorganize and improve higher education in Australia.						

References: 1. *Research and Experimental Development, Higher Education Organisations* (Australian Bureau of Statistics, 2010), available at go.nature.com/1obhsa; 2. United Nations Educational Scientific and Cultural Organization Institute for Statistics; 3. Kwok, J. T. *Impact of ERA Research Assessment on University Behaviour and their Staff* (NTEU National Policy and Research Unit, 2013) available at go.nature.com/c2ypxa;

2010

All data are from 2010 unless otherwise stated, because 2010 provides the most recent, reliable, comparable and complete data. All dollar amounts are in US dollars, converted using a 2010 exchange rate.

NEW ZEALAND RESEARCH ASSESSMENT FUNDING BY UNIVERSITY

The majority of New Zealand's higher education funding is allocated by the Performance-Based Research Fund (PBRF), which spent a total of $168 million on tertiary education organizations in 2010.

Population:

4.4M

Auckland University of Technology **$6M**

University of Auckland **$56M**

• Auckland

University of Waikato **$12M**

University of Canterbury **$21M**

• Wellington

Massey University **$27M**

Lincoln University **$7M**

• Christchurch

$18M Victoria University of Wellington

University of Otago **$41M**

• Dunedin

Funding for each university allocated by the PBRF[4].
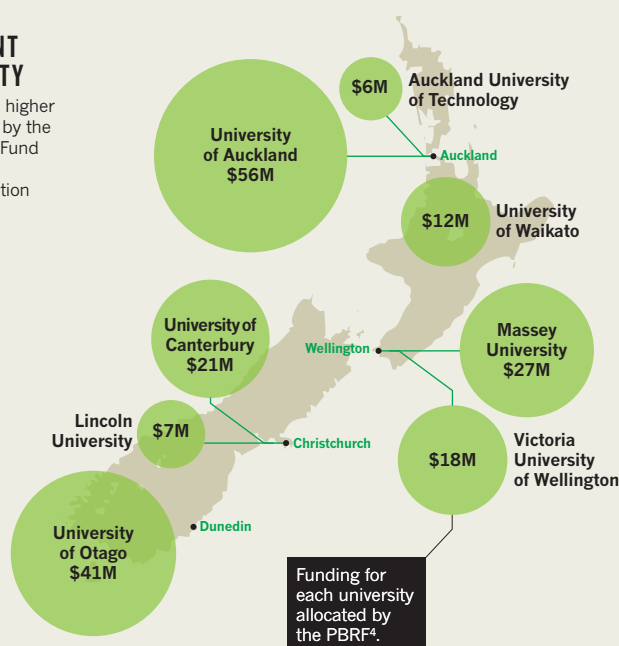
The Performance-Based Research Fund defines research as:

"Original investigation undertaken in order to contribute to knowledge and understanding and, in the case of some disciplines, cultural innovation or aesthetic refinement."

THE GREAT DIVIDE

The system used by the PBRF to allocate funds to universities[4].

100%

60% — Quality evaluation

25% — Research degree completions

15% — External research income

*The system used to allocate funds by Excellence in Research for Australia is more complicated and relies on a more peer-review structure than New Zealand's system.

PRINTED MATERIAL

Any published articles* that include at least one author from Australia or New Zealand[5].

Number of published articles

Australia

New Zealand

Year: 2000 2002 2004 2006 2008 2010
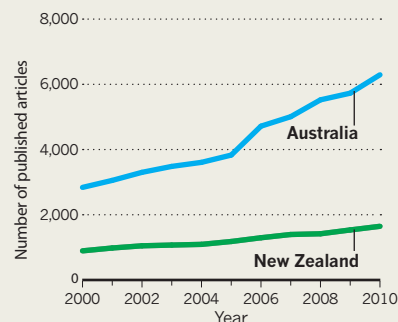
*Journals, conference proceedings, book series, books or trade publications.

MONEY MATTERS

Almost 90% of all funding for higher-education research and development (HERD) in New Zealand is provided by the government and universities[6].

Other **$21M**
Overseas **$19M**
Business **$32M**

$0.6B Sources of funding

Government (inc. PBRF) **$399M**

Universities (inc. student fees) **$147M**

Experimental development **$100M**

$0.6B HERD expenditure

Basic research **$298M**

Applied research **$219M**
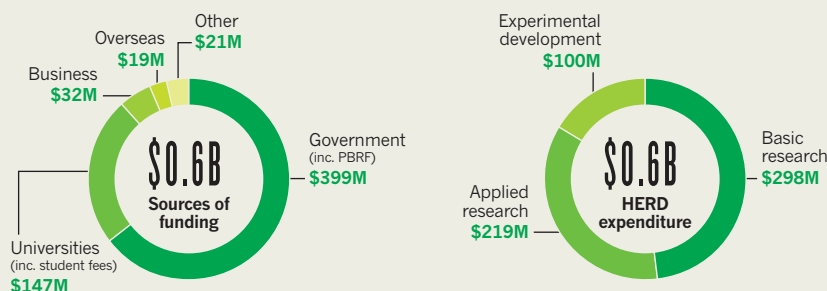
PERFORMANCE-BASED RESEARCH FUND: TIMELINE

Before 2000, funding was provided to tertiary education organizations based on the number of researchers. Now funding is dependent upon individual performance reviews[7].

1990	1995	2000	2005	2010	2015
Education Act establishes the Tertiary Education Commission.	White Paper to assure research quality and accountability.		First Performance-Based Research Fund (PBRF) assessment.	PBRF reviewed. All funding now allocated through the PBRF.	Full PBRF assessment.
	Funding now based on number of equivalent full-time students, weighted by course costs.	Tertiary Education Advisory Commission (TEAC) established to provide advice on strategic direction.	TEAC introduces the PBRF.		Next PBRF assessment is in 2018.

4. *Performance-Based Research Fund Annual Report 2010* (Tertiary Education Commission), available at go.nature.com/s5t6p7; 5. Scopus; 6. *Research and Development in New Zealand 2010* (Statistics New Zealand), available at go.nature.com/c7bbp5; 7. *A History and Overview of the PRBF* (New Zealand Ministry of Education), available at go.nature.com/vwXjbu.

SDP MEDIA

Multidisciplinary funding and the benefits of research evaluation programmes were hotly debated at an event organized by *Nature*.

Calling science to account

Systematic evaluation of scientific research might strengthen public support, but could it also stifle innovation? The issues were debated at a symposium in Melbourne.

BY TIM THWAITES

Since the United Kingdom's first Research Assessment Exercise in 1986, the concept of a national evaluation of publicly funded research has expanded to other countries, including Belgium, France, Italy, Australia and New Zealand. Some assessments are performed specifically to determine allocation of research funds, whereas others are benchmarking exercises of the performance of local research in a global context. Although the overall goals of these assessment systems are well understood, there is doubt as to how well each is working.

Their relative effectiveness was the focus of a symposium in February 2014 in Melbourne, Australia. *Nature* brought together experts from institutes and universities in Australia, New Zealand and Singapore to examine issues surrounding the outcomes and impact of how research is measured.

In his introduction to the symposium, *Nature* editor-in-chief Phil Campbell outlined several of the issues and views that were later discussed. "There is a need for research evaluators to be explicit about the methods they use to measure impact," he said. "Openness is an essential part of earning trust. *Nature* welcomes a diversity of indicators." Relying solely on citations, Campbell added, "absolutely can't be sustained".

The United Kingdom has recently reoriented its research-assessment programme to bring peer review, case histories and metrics into a system called the Research Excellence Framework (REF), which runs for the first time this year (see 'How research benefits the United Kingdom'). The Melbourne symposium examined this approach against various schemes in the Asia-Pacific region, including Singapore's carefully programmed development of knowledge-based industry; New Zealand's proposition that criteria for assessment be laid down

even before research starts; and Australia's quantitative evaluation of its research strengths and weaknesses. Two things were clear: there are many reasons for evaluating research, and there are lots of approaches to get results. Perhaps the first hurdle to overcome is deciding what you want to achieve.

VALUING RESEARCH

The symposium's keynote speaker was David Sweeney, the director for research, innovation and skills at the Higher Education Funding Council for England (HEFCE) in Bristol. Sweeney, who managed development of the REF, told delegates there was "no right to research funding". He said, "If, as happened in previous budget proposals in the UK, senior scientists say to government 'Give us the money, and we will deliver the goods', the treasury has a right to say, 'Prove it!'."

Sweeney said that scientists cannot assume that the general public understands the value

of their research, so evaluation has become an essential tool for convincing UK government, business and society why they should invest in universities and research. In fact, he said, the UK government wanted to enlist companies to help fund university research — unlocking some of the capital that businesses had put away during the global financial crisis to protect against hard times. The outcomes of the REF, teamed with matched-funding schemes, could help the government release previously hidden private pots of money, he argued.

Sweeney outlined the REF's methodology. "Academic excellence is still the number one objective of public funding," he said. But conventional gauges of merit, such as peer review and citations, should not comprise the whole assessment; it's also important, he said, to reward research that has a positive impact on society. He asserted that the REF did not open the way for government to dictate research direction. Nor did it mean a bias towards funding applied research. Instead, said Sweeney, REF provided a means of validating the contribution of all research: "It's not about favouring one discipline over another." He presented REF not as a perfect measure of impact, but as a first step. "The methodology does the job that needs to be done now, even though it's not perfect," he explained.

Real-world issues, such as water and energy usage, are complex and interconnected, and research addressing these matters needs to draw on expertise from physical and biological sciences, as well as social sciences including economics, behavioural psychology and law. Yet, according to participants in a panel discussion on multidisciplinary research, such crucial work has rarely been valued appropriately in research assessment exercises.

The intrinsic value of multidisciplinary teams, and the difficulties of their coordination, were well illustrated by the story of the Murray-Darling Basin Plan, set up to manage water resources in Australia's largest and most agriculturally productive area. "It was a wonderful document that told us exactly what we should do," said Robert Saint, pro vice-chancellor of research strategy at the University of Adelaide. The plan was unpopular as it proposed swingeing cuts to water allocation for many farmers. "Its release was closely followed by farmers burning it, and the whole business had to go back to the drawing board." The problem was that the Murray-Darling Basin Authority, which compiled the report, lacked the specific capabilities for incorporating legal, political and social issues alongside the science.

Australia's largest national research body, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), based in Canberra, is no stranger to multidisciplinary research, said its chief executive Megan Clark. CSIRO, she noted, specializes in large-scale, broad, "pan-disciplinary" research groups. "There is an understanding from the minute you walk in that this is not a place to work on personal research," she said. "We work in multidisciplinary teams on mission-directed research." As a result, CSIRO's evaluation of its own research includes traditional outputs, such as patents and journal publications, and quality assessments by independent peer review panels, but crucially also takes into account the impact of its work on end users — including the public, government departments, private companies and environmental organizations (see page S72).

"The methodology does the job that needs to be done now, even though it's not perfect."

CSIRO runs large-scale multidisciplinary research partnerships known as National Research Flagships (see 'Launching flagships'). These focus on issues of national and global importance such as biosecurity, preventative health, manufacturing and sustainable agriculture. In a little more than a decade, the Flagships programme has grown to encompass more than half of all CSIRO research activity.

Many stakeholders, Clark recalled, feared that the Flagships programme would lead to a decline in the quality of the organization's science. But CSIRO's experience has been the reverse, she said. "Last year, we hit a record in the quality of our science and our standing globally." For instance, the citation rate for CSIRO research publications is now 56% more than the global average, according to the organization's latest *Science Health and Excellence* report.

CSIRO's approach differs from multidisciplinary work undertaken at universities, which are the primary training grounds for researchers, said Kim Langfield-Smith, vice-provost for academic performance at Monash University in Melbourne. The academic environment tends to have discipline-focused organization underpinning promotion tracks. This silo structure is not conducive to researchers thinking outside their speciality.

Langfield-Smith spoke of the difficulties in recruiting university researchers for multidisciplinary projects. In particular, mid-career and older researchers found it difficult to justify interrupting their research to join projects that might not yield publications in the top journals of their own fields. What's more, multidisciplinary research is difficult to get underway: it routinely lacks common language, modes of analysis, conceptual frameworks and dedicated journals (many outcomes

---

HOW RESEARCH BENEFITS THE UNITED KINGDOM
Evaluating research can increase its public support

The Research Excellence Framework (REF) is the new system for assessing the quality of research at UK universities. It aims to demonstrate the benefits of public investment in research; to show accountability for government research funding; and to rate the quality of the United Kingdoms's research efforts on a global scale across all academic disciplines.

The REF — successor to the Research Assessment Exercise — will produce its first report in December 2014. It will be used to assist the four UK higher-education funding bodies — the Higher Education Funding Council for England, the Scottish Funding Council, the Higher Education Funding Council for Wales, and the Department of Education and Learning in Northern Ireland — to allocate government funding for research, at present about £1.6 billion (US$2.7 billion) a year.

It assesses the efforts of higher education institutions across 36 subject areas, determined by the quality and impact of research as well as what the research environment is like (see 'UK scorecard'). The results will be published in the form of a quality profile in which each submission is ranked as either world-leading,

UK SCORECARD
How a Research Excellence Framework (REF) score is determined.

- 15% Research environment
- 65% Research quality
- 20% Impact beyond academia
- Total REF score
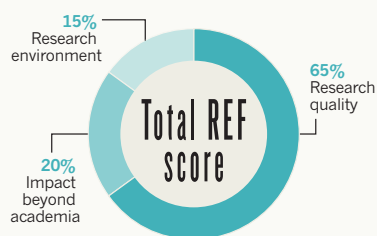
Research quality (up to four papers per researcher, published between 2008 and 2013).
Impact beyond academia (case studies written by researchers).
Research environment (research strategy, facilities, staffing, etc.).

internationally excellent, internationally recognized, nationally recognized or unclassified.

The assessment is made by expert panels, which include representation from people in business or government who use research outputs in their professional activities or who commission or collaborate with academic researchers. **T.T.**

are instead published as government reports).

Saint observed that peer review could be disadvantageous to multidisciplinary projects at both the funding and publication stage. "I remember the early days of bioinformatics: statisticians would argue that all the theory had been done 40 years ago, and biologists couldn't see anything interesting in statistics." The panellists suggested several ways to promote multidisciplinary work, including setting up dedicated funding streams for such research, and altering the criteria of assessment so that work published in government reports is eligible for consideration.

Hugh Durrant-Whyte, chief executive of NICTA, Australia's largest information and communications technology research organization, suggested that the solution lay in removing disincentives. Funding agencies, he said, should foster a research culture that encourages scientists to undertake projects because they were "cool and exciting, not because there is a paper at the end". Young scientists, he said, should be urged to "find something interesting and get on with it". This would naturally stimulate collaborations and multidisciplinary work, he added.

WHEN OPTIONS ARE LIMITED
The value of research to government can be very different from its value to business, or to academia or the public. That's why it's critical to set the criteria for evaluation from the very beginning, said Peter Gluckman, chief science adviser to the prime minister of New Zealand. This approach "changes the way research is done", said Gluckman. "It influences how scientists work and think."

Gluckman was mainly referring to government-directed projects that account for a large portion of the science budget of small countries such as New Zealand. Perhaps the most compelling argument for this principle can be seen in Singapore, which has taken little more than a decade to generate a biomedical industry from a low starting point (see 'How to grow an industry').

David Lane is chief scientist of Singapore's Agency for Science, Technology and Research (A*STAR) which, with the country's Economic Development Board, was responsible for implementing the Biomedical Sciences Initiative to develop the industry. He said that determining impact was a major part of the government's strategy. "Our budget was increased," he said, "but 30 to 40% was set aside and would only be released if we could show we were doing work aligned with industry." The yardstick by which the effort was measured was the level of corporate investment in Singapore's biomedical industry.

Such a utilitarian view of science by governments, said Gluckman, differs enormously from the academic perspective, which focuses on accumulation of knowledge. Governments, he said, were typically concerned with research

impact on the economy, the environment, defence and public health. Such priorities were greater in small economies that cannot so easily spare money for blue-sky research.

One of the purposes of the Excellence in Research for Australia (ERA) programme at its inception was to determine in which research fields Australia had world standing, said Margaret Sheil, provost of the University of Melbourne and a former head of the Australian Research Council (ARC). Sheil, who was heavily involved in the design and operation of the ERA (see page S67), pointed out that although Australia had a small population, it was competing globally in many disciplines.

Representing these different viewpoints in one assessment tool is not easy. Science

entrepreneur and chancellor of Monash University, Alan Finkel, suggested that funding bodies needed a framework where activities such as working in industry, contributions to government reports or communication of research outcomes to audiences other than a researcher's peers could be converted into a "citation equivalent" for the purpose of improving the measurement of research impact (see page S77).

METRICS ARE NOT THE ANSWER
Assessing a country's research enterprise is not an end in itself. And when it comes to acting upon the outcomes of research assessment, funders have vastly differing viewpoints. The one issue on which they tend to agree is

---

LAUNCHING FLAGSHIPS
Restructuring CSIRO to tackle national problems

The National Research Flagships are large-scale, multidisciplinary research partnerships between the Commonweath Scientific and Industrial Research Organisation (CSIRO; Australia's largest national research body), universities, other publicly funded research institutions, the private sector and international organizations. There are 11 current Flagships tackling significant national challenges (see table, below).

The initiative began with three Flagships in 2003 and has grown into one of largest scientific research programmes ever undertaken in Australia. Together, they

account for more than half of CSIRO's budget of about AUS$1.5 billion (US$1.4 billion), expected to increase to 65% by July 2015.

Setting up the Flagships required a substantial organizational shake-up. Initially, CSIRO retained its 11 traditional discipline-based divisions across the country, with the Flagships able to draw staff from any division or partner organization. From July 2014 they will merge into nine Flagships.

Since 2005, a Flagship Collaboration Fund has committed more than AUS$130 million in grants to encourage partnerships between universities, CSIRO and other research agencies. **T.T.**

FLAGSHIP TITLE	YEAR LAUNCHED	DESCRIPTION
Energy	2003	Investigating low carbon alternative energy sources and the future of transport.
Food Futures	2003	Transforming the global reach of the Australian agrifood sector.
Preventative Health	2003	Improving well-being through research into prevention, detection and health interventions.
Wealth from Oceans	2003	Investigating the network of resources, capacity and sustainability of the oceans and developing appropriate technologies.
Water for a Healthy Country	2004	Developing technologies to improve the social, economic and environmental outcomes around water access and use.
Future Manufacturing	2007	Developing cleaner advanced materials and technologies.
Minerals Down Under	2007	Growing Australia's resource base, increasing productivity of the minerals industry and reducing its environmental footprint.
Climate Adaptation	2008	Supporting Australia's efforts to adapt to climate change.
Sustainable Agriculture	2010	Addressing productivity and food security in a carbon-constrained world.
Digital Productivity & Services	2012	Developing and delivering improved online services and changing the way people engage with technology.
Biosecurity	2013	Helping to protect Australia from biological risks posed by exotic and endemic pests and diseases.

HOW TO GROW AN INDUSTRY

Singapore becomes a biomedical powerhouse

Singapore's impressive bioscience research hub, Biopolis, which opened in 2003.

In little more than a decade, Singapore has established a thriving biomedical industry from scratch. The country, with few natural resources, set itself a goal in the early 1990s to become a knowledge-based, innovation-driven economy. The government identified biomedical sciences as an area with tremendous growth potential and decided to try to grow its own industry. In 2000, it started to invest in a Biomedical Sciences initiative and, by 2012, this industry had grown to more than SG$30 billion (US$24 billion), comprising more than 50 manufacturing plants, 50 R&D centres and 30 regional headquarters of biotechnology and pharmaceutical companies.

The first phase of the initiative (2000–05) concentrated on rapidly expanding Singapore's basic research capabilities and infrastructure as well as attracting industry R&D laboratories from other countries. The government also constructed an R&D hub, the Biopolis — seven buildings of 185,000 square metres at a cost of SG$500 million — next to the National University of Singapore.

The second phase (2006–10) bolstered Singapore's capacity to undertake translational and clinical research and turn it into health-care products and applications. Large biomedical corporations, such as Lonza and Genentech, were attracted by favourable business and investment terms including generous tax rates and access to a flexible and skilled labour market.

The current phase (2011–15) focuses on encouraging international investment and links with industry. An Industry Partnership Office has been set up to facilitate collaborations with private enterprise. The five-year budget was increased by 16% to nearly SG$16 billion, with a substantial portion channelled into industry-oriented research. **T.T.**

Research Council (NHMRC) in Canberra, which dispenses more than AUS$750 million (US$700 million) of government money in research grants each year (see page S52).

"It's not only the research that's important, but also how it is used," he said. Health researchers typically wish to influence decision-makers and medical practitioners as well as other scientists, which means they need to publish in areas outside academic literature. To properly evaluate their work, he said, you needed to consult sources other than scientific journals, such as government reports and health-care experts. Government has a huge interest in health care because of its enormous cost. Australia's AUS$140 billion health-care industry — comprising vaccine manufacturers and medical device developers, among others — is also the nation's second largest exporter of manufactured goods, Anderson said.

Australia's other major research funding body is the ARC, responsible for disbursing more than AUS$900 million a year. It also administers the ERA, which aims to determine areas of Australia's research strengths. ERA assessments are made by internationally recognized researchers, organized by discipline and clustered into eight Research Evaluation Committees. They use traditional measures of quality, such as citation analysis or peer review, but also incorporate a broader view, considering income from commercialization and measures of esteem — for example being admitted to a learned society such as becoming a fellow of the Australian Academy of Science.

ERA ranks research quality against a global scale and is "a rigorous and robust measure across all discipline domains", ARC's chief executive Aidan Byrne told the symposium. It aims, he said, to get researchers to change their focus from quantity of work to quality. "In that, the ERA exercise has been spectacularly successful. And it did it without tying the exercise to financial rewards."

Furthermore, despite its reliance on metrics, ERA results for academic excellence correlate with other real-world outputs, Byrne said. For instance, 95% of industry investment in research in Australia is in the same areas in which researchers performed at world-class or better. And the same is true for 98% of the research that was commercialized and for 97% of the work that was patented.

HEFCE's Sweeney's take on various methods of assessment was straightforward. No system will be perfect, he said, but you have to start somewhere: "You can propose alternatives, and spend five years discussing them, but that's not going to solve today's problems." ∎

Tim Thwaites *is chief science writer for Science in Public, a science communication agency based in Melbourne.*

that any worthwhile evaluation of research — whether it be for disbursing grants or encouraging excellence — needs to be based on a range of measures, not just the quantity of publications and how often they are cited by others. In the final panel of Melbourne symposium, representatives of four significant funding organizations discussed how best to incorporate the information gained from assessments.

Traditionally, research assessment evaluates completed projects. But, in an ever-changing research environment, a scientist's past successes might not be a predictor of how well they will perform in the future, said Tony Peacock, chief executive of the Australian Cooperative Research Centre (CRC) Association in Canberra, which runs the nation's 40 CRCs — collaborative partnerships between publicly funded researchers and industry. In fact, Peacock argued, rewarding only those strategies that were successful in the past would tend to discourage new approaches and stifle innovation, the essence of successful science. Relying solely on citation and peer review metrics was opposed for similar reasons by Warwick Anderson, chief executive of Australia's National Health and Medical

RESEARCH ASSESSMENT

The limits of excellence

Young researchers and interdisciplinary science might be getting short-changed by research assessment in Australia and New Zealand.

BY ANNABEL MCGILVRAY

The day before he and I speak, Jonathan Boston receives an email on a familiar theme. It is from a colleague concerned about a junior researcher whose career decisions are "being twisted in an uncomfortable way" by the demands of New Zealand's Performance-Based Research Fund (PBRF). "I have had many such messages over the years — which reflect the good and the bad of the PBRF," says Boston.

Boston holds a personal chair in public policy at Victoria University of Wellington, and was one of the architects of the New Zealand system. He says that research paths can become conflicted by one of three scenarios: compulsion to publish articles in

high-impact international journals rather than working on a book with a domestic publisher; pressure to change research focus to better align with mainstream or more highly esteemed fields; or encouragement to accept a position as a non-PBRF-eligible teaching fellow and move away from active research. Each of these outcomes can be traced to the way in which the PBRF measures research excellence, and so its influence on the country's research funding environment.

At its conception in 1999, explains Boston, the PBRF was an ambitious undertaking to measure research excellence and raise standards at institutions across New Zealand. It was intended to remedy years of neglect of the research sector through which the bulk of funds, dubbed research top-ups,

had been linked to postgraduate student numbers. The flaws of that arrangement had become evident in the mid-1990s when non-university higher education providers — such as polytechnics and institutes of technology — began offering postgraduate degrees. The research pot was suddenly being split between ever more institutions, many of which had limited research capacity. When Helen Clark's Labour government came to power in 1999 with five former academics — including Clark, a lecturer in political studies — among its senior ranks, it vowed to strengthen the process of research funding and increase accountability. "The only option was some sort of performance-based regime," says Boston.

The PBRF is based on the individual, making it unique among measures of national

research excellence. Every six years, it gauges and reports the standard of research of each of New Zealand's approximately 6,000 researchers in universities and colleges (so-called tertiary educational establishments). These rankings — A, B, C and R — are provided to the institutions; a researcher can apply to receive his or her own rating. The outcomes are then weighted by quality and subject area, in line with the resources required for different fields. The individual results are aggregated by institution and are the major determinant, alongside external income and research degree completions, of the distribution of research funding. The PBRF is now the largest single source of tertiary research funding in New Zealand, worth NZ$262.5 million (US$224.2 million) in 2013 (see page S52).

Across the Tasman Sea, the Australian Research Council (ARC) is gearing up for the third round of its own national measure of research quality, the Excellence in Research for Australia (ERA) evaluation, which takes place for every three years for the country's 41 universities. The forthcoming ERA2015 will categorize and evaluate the nation's entire higher education research output, comprising more than 400,000 publications. The ERA assesses work by discipline, and the initiative directly influences only a small portion of university research funding.

Results suggest that the overall quality of research has increased in both countries since the introduction of national assessments. More New Zealand researchers are achieving an A rating, and more Australian disciplines are classed at 'above world-standard'[1,2]. But the PBRF and ERA prompt passionate reaction in their respective research communities. With increasing awareness of the need to assess the societal impact of research, merely weighing academic excellence makes less sense. There is concern that subject-focused assessment programmes don't adequately recognize the value of interdisciplinary research. And, as Boston's recent email correspondence implied, there are fears that the way excellence is measured — in particular the focus on high impact publications — may be hindering the careers of young researchers.

THE FOLLY OF YOUTH
PBRF and ERA both use metrics of quality and peer review to determine ratings. The same indicators are used in different forms in research assessment schemes around the world. However, for the PBRF, all nominated publications and other research outputs are rated by selected reviewers — a qualitative process that depends on individual judgements.

The ERA, by contrast, places more emphasis on citation analysis. "The number of people who are citing and making reference to work is a pretty good indicator of the significance and impact it has had in the academic community," says Aidan Byrne, ARC chief executive officer.

But these measurements of excellence are creating obstacles for young scientists, says Attila Brungs, deputy vice-chancellor for research at the University of Technology, Sydney (UTS). "Narrow metrics can drive some bizarre behaviours. People don't publish as much with PhD students because PhD students are often published in lower-ranked journals."

More broadly, the strengths of early-career researchers aren't readily demonstrated by reference to an objective publication review, a particular flaw of the individual-centred PBRF. Assessment encourages institutions to employ staff with established research records rather than emerging researchers who are doing excellent science but who are yet to amass publications. There is early evidence of this reluctance to engage young researchers, with one study showing a 14% drop in research staff aged 35 and younger between the first and second rounds of the PBRF[3].

There are fears that the way excellence is measured may hinder the careers of young researchers.

Indeed, a 2008 independent review of PBRF found that morale of otherwise high-achieving young researchers was being hurt by low ratings. The review, commissioned by the body that oversees PBRF, the Tertiary Education Commission (TEC), stated that "the assignment of a 'C' grade was seen by rising stars to undermine morale and to stigmatize their position". Boston says that when the PBRF scheme was designed there was no intention to reveal individuals' ratings. Not until after the system was established did Boston and his colleagues realize they were compelled to impart that information to researchers. "We simply failed to fully realize the implications of the Privacy Act and Official Information Act," he says. "If I had known we would end up with a regime in which individuals had their scores reported to them, and that other people could potentially know what they were, I would not have supported it."

The TEC has created a specific 'new and emerging researcher' category to counter disincentives to employ early career researchers when evaluation rounds loom. Researchers in this category can qualify for C(NE) rating and contribute to their institution's funding allocation. Their evidence portfolio assessment is weighted against their time as a researcher, with a minimum of two research outputs generally expected.

But many people, including Boston and Peter Gluckman, chief science advisor to the New Zealand prime minister, believe that the individual judgements inherent in the PBRF reviewing process continue to place undue pressure on emerging researchers to publish in high-impact journals. "I've seen several young

researchers quite compromised by this drive to produce the one paper that will get into *Nature*," says Gluckman, "when their career would have been much more developed had they focused on getting solid, excellent papers in the appropriate journals".

TOGETHER YET APART
Researchers undertaking interdisciplinary work are also feeling compromised. Campuses across New Zealand and Australia are bringing together researchers from multiple disciplines, from the hard sciences to the humanities, to look at societal problems in a holistic way. These fields include environmental sustainability and medical research and, in many cases, work is carried out under the auspices of a centre or an institute within a university.

Despite this big-picture approach, assessments such as the PBRF and the ERA continue to view research through a mono-disciplinary lens. The final report from the 2012 PBRF conceded that the 42 subject areas under which all research is assessed "do not accurately reflect the way research activity is organized and conducted". Despite this acknowledgment, there are no plans for a review, says Marny Dickson, chief policy analyst for tertiary education at the New Zealand Ministry of Education.

The story isn't much more encouraging in Australia. The Australian Council of the Learned Academies (ACOLA) — representing the four Australian academic societies: the Australian Academy of Science, the Academy of Social Sciences in Australia, the Australian Academy of the Humanities and the Australian Academy of Technological Sciences and Engineering — seeks to inform policy specifically related to multidisciplinary research. In a 2012 report, ACOLA found that the ERA "has difficulty in evaluating and reporting interdisciplinary research". And the situation is likely to be exacerbated as universities base their internal benchmarks around the ERA, which, like the PBRF, focuses researchers on higher impact journals — few of which are interdisciplinary.

For instance, the Centre for Cosmopolitan and Civil Societies at UTS, does a lot of applied research related to policy, and frequently produces work for the local, state and federal governments. When the centre's researchers publish, they have to do so in the journals of their individual expertise, whether marketing, business and economics, or social science and the humanities. The university is then rated separately for each of these fields, rather than for the centre's projects as a whole. "In an exercise like ERA, their work disappears, it doesn't exist," says Brungs.

Brungs says that the distortion doesn't yet affect the university's research priorities, because the funding linked to ERA is very small. In 2014 it was just AUS$69 million, or 4% of the research block grants made by the Department of Education. But, an increase in

INTERNALIZING TARGETS

Measurements used by the Excellence in Research for Australia system are finding their way into internal targets set by research institutions. Here is one example of 2011 targets from a leading Australian university.

Research outputs	Laboratory-based sciences: minimum (Min) and aspirational (Asp) targets							
	Lecturer		Senior lecturer		Associate professor		Professor	
	Min	Asp	Min	Asp	Min	Asp	Min	Asp
N° publications	1.5	4	3	6	4	8	7.5	15
Impact factor	4.5	16	9	24	12	32	22.5	60
Proportion in A/A* journals †	35%	55%	45%	65%	55%	80%	55%	80%
Research income (AUS$)	$5,000	$40,000	$30,000	$125,000	$150,000	$500,000	$500,000	$1,000,000

† Ranking of journals as a proxy for quality (A/A* being the two highest) was abandoned by the ERA in 2011.

the proportion of ERA funding would make future collaborations harder to justify. "Universities are not allowing the drivers to distort their behaviour too much," says Brungs. "But if we continue to go down that path it does have real danger for the interdisciplinary sector."

PLAYING GAMES

If PBRF and ERA become more significant, recruitment policies at universities will inevitably be coloured by how a candidate might affect a pending evaluation. Some fear this will lead to a widespread gaming of both systems as institutions try to improve their scores.

Australia's National Tertiary Education Union has found that ERA gaming already occurs, as individuals, departments and institutions strive for results needed to influence funding decisions that last for three years. In the case of the PBRF, the equivalent decision influences six years of funding, a magnitude which further incentivises manipulation. The problem is likely to persist, says Frank Larkins, former deputy vice-chancellor for research at the University of Melbourne. "Universities have a lot of smart people and they can learn pretty fast how to optimize their performance," says Larkins.

The ERA peer-review panels are asked to look closely for idiosyncrasies in the research performance of institutions, and the ARC is now able to cross-reference dubious submissions against previous rounds. However, there is nothing to stop universities taking on researchers, and sometimes whole research departments, in order to boost output prior to an ERA round. The Australian newspaper described the "churning" of researchers this year in the lead-up to the 31 March 2014 census deadline. Any staff hired after this date are not eligible for assessment in ERA 2015. But, during the preceding Australian summer, research groups and even whole departments were poached by the Australian Catholic University, Central Queensland University and Charles Sturt University, among others.

ARC leader, Byrne, doesn't endorse such activities, but says that calculated reallocation of resources for the purposes of ERA ranking is not necessarily a bad thing. "We don't want to stop institutions from making strategic decisions about what research they wish to pursue." The importance placed on the ERA rankings by Australian university management has been evident, not only in the tendency for researcher churning, but also in the tailoring of internal research benchmarks to better meet the terms of the ARC system (see 'Internalizing targets'). Consequently, universities have set departmental and school-wide targets regarding quantity and quality of publications.

Despite this, Byrne does not accept that ERA is forcing institutional change. "It does get used by institutions in various ways, but we are providing an evaluation against the best possible standards we can come up with", he says.

Perhaps the most egregious example of an attempt to game the system occurred in New Zealand in 2006. One leading university reclassified dozens of staff members, notably those who were PBRF-eligible but performed little active research. By reclassifying inactive researchers away from subjects such as economics and biology to fields such as philosophy and religious studies, the university would improve its standing in the former fields. The surge in the number of New Zealand philosophers piqued the curiosity of PBRF reviewers who eventually reversed the classifications.

MEASURING A MOVING TARGET

The objectives and the structure of ERA and PBRF have changed little, but the status quo may be threatened by demand for the explicit inclusion of research impact as a quality indicator within the assessment exercises (see page S81). At UTS, Brungs says that more focus on impact might bring much-needed formal recognition of interdisciplinary work within the system. For example, work that has valuable outputs concerned more with policy than scholarship. "Publishing in *Nature* is one way of demonstrating excellence in research," he says. "Changing the way that a nation drinks water is another way."

The ARC is considering the inclusion of impact measurements, but Byrne says the organization does not want to just graft these on to the existing system and does not have the resources to develop an independent measurement of impact. The Australian government's current aversion to any increase in red tape does not help.

In New Zealand, the 2008 review of PBRF cautioned against diluting its focus on excellence by aligning it with government innovation policy. However, a re-evaluation has seen a number of alterations, including an increase in the significance of investment from industry in determining overall funding awards, coupled with a moderate reduction on the emphasis of the research quality assessment. According to the Ministry of Education, these changes reflect the fact that external research income is a "strong proxy indicator" for the transfer of knowledge between academia and industry and the change will encourage "research of relevance to end-users". It is a tangible shift towards reward for research impact.

But Boston is not satisfied that, even with such changes, assessments like the PBRF and ERA will continue to be relevant. He refers to Goodhart's law, which states that once a measure becomes a target it ceases to be a good measure. "I don't see the logic of running the same assessment process every six years ad infinitum, with only minor tweaks," he says. "It sets up a particular set of incentives and a particular kind of process within institutions, some of which is undesirable."

To keep improving research excellence, Boston says, the government needs to increase funding and other resources, or make bigger changes to the assessments, for instance by introducing new criteria. Continual re-allocation of finite resources can only do so much. "I don't know how you're going to squeeze more drips out of the orange." ∎

Annabel McGilvray is a freelance science and medical writer based in Sydney.

1. *Performance-Based Research Fund: Evaluating Research Excellence — the 2012 Assessment, Final Report* (Tertiary Education Commission, 2013); available at go.nature.com/gprpet.
2. *Excellence in Research for Australia 2012, National Report* (Australian Research Council, Commonwealth of Australia, 2012); available at go.nature.com/q8sxiw.
3. Çinlar, N. & Dowse, J. *Staffing and Performance Trends in Research Subject Areas* (Tertiary Education Commission, 2008).

PERSPECTIVE

On the verge of a new ERA

Despite its limitations, Excellence in Research for Australia was the right assessment tool at the right time, says **Margaret Sheil**.

In the early part of this century, the Australian government made a decision to improve the country's research effort. It made targeted investments in research agencies, built infrastructure and created a number of world-class centres of excellence. Yet some argued our research efforts were spread too thinly and performance was uneven. By 2007, only two of Australia's 40 universities were in the top 100 of the Academic Ranking of World Universities (ARWU). That same year, an Australian Productivity Commission report on public support for science and innovation identified several areas of the innovation system requiring improvement, which included a lack of effective support for industry-based research and development, deficiencies in the scientific workforce and inadequate methods of evaluation.

The government determined that Australia's dual funding system, comprising individual competitive grants and block grants to institutions, was not providing the right incentives. In particular, block funding was distributed according to formulae that took into account only the number of scholarly publications, with no consideration for their quality. The results were, in hindsight, predictable: the number of publications increased but not their overall quality. The government needed a new framework: one that would encourage universities to focus their endeavours and build on their strengths; to address weaknesses in research performance; and to provide researchers with an incentive to target quality. All of that had to happen within a system that was streamlined and cost effective.

CORE VALUES

From 2007 until 2012, I was chief executive of the Australian Research Council (ARC), the body tasked with developing the new framework. We weren't starting from scratch; the previous government spent nearly four years working on a new design, and we could draw inspiration from other assessment systems, including the United Kingdom's Research Assessment Exercise (RAE), and New Zealand's Performance-Based Research Fund (PBRF).

None of the approaches, however, were entirely appropriate for Australia. The proposed design from the last government relied on case studies to assess research impact. Case studies are expensive and time-consuming to prepare and assess, yet were not, the new minister felt, sufficiently robust to inform funding allocations. Similarly both PBRF and RAE — to different degrees — are selective exercises, focusing on just the best output, allowing universities to hide poor performance. RAE used panels of experts to assess the quality of publications — a costly and lengthy process. And the units of assessment for both schemes were unsuited to Australia: PBRF evaluates individual portfolios, making it difficult to scale up; RAE evaluates departments. But our universities have varied and complex organizations of departments, schools, faculties and/or research centres which do not lend themselves to a simple comparison.

These schemes did, however, provide useful leads. In particular, the 21 years of the RAE saw a growing correlation between the quality ratings assigned to departments and their citation performance. This suggested that metrics alone could be used as indicators for many scientific disciplines, although less so for others, such as the humanities and mathematics where citation data were less reliable or books were more important than journal articles.

In coming up with our own solution, the ARC consulted widely with institutions, learned academies, research leaders and bibliometric experts. We settled on 'discipline' as the best unit for evaluation because it avoided focusing on either university structures or individuals (thereby minimizing the value and prospect of poaching individual stars). Where citation analyses were not appropriate, peer review of selected publications or outputs was used instead. This approach also had the benefit of enabling evaluation of the creative arts — important for Australia where most conservatoria and visual and performing arts schools had been incorporated into universities.

CITATION ANALYSIS CAN BE USED AS A PRINCIPLE INDICATOR OF QUALITY.

Excellence in Research for Australia (ERA) was born in 2010. By including all outputs within a discipline, rather than just the premier efforts of a select group of researchers, ERA ensured that the attention of every researcher was on quality rather than quantity. And because its discipline-specific measures of achievement accorded with established academic practice, the results were largely accepted with only a small number of disputed outcomes.

There have been difficulties. ERA initially ranked around 22,000 journals in four bands (A*, A, B and C) to provide another set of indicators for the committees to use. However, some institutions used the journal ranks out of context, potentially harming the careers of young academics and those working in cross-disciplinary areas, since both groups are unlikely to publish in top tier journals. Bandings were removed in 2012.

Overall, ERA has been of considerable value — despite the fact that it drives only a small proportion of the block funding allocations. Crucially, it has demonstrated that citation analysis can be used as a principle indicator of quality in many disciplines, producing finely grained information about research strengths and weaknesses. Governments, universities, industry and the academic disciplines themselves have welcomed this information and make regular use of it. This success shows in our international standing: Australia now has 5 universities in the ARWU top 100, and 19 in the top 500.

Any comprehensive evaluation system is bound to have its critics. But widespread consultation and the use of discipline-based solutions has helped minimize opposition to the ERA. Though there is widespread acceptance that there are benefits in competitive processes for teams and individuals that have winners and losers, the political dimensions associated with 'losing' loom larger when applied to institutions, for example, negatively affecting newer institutions in areas serving growing and diverse populations. These considerations are beyond the control of the ARC and are unrelated to the ERA itself, which, while not perfect, has nonetheless achieved its goals. ■

Margaret Sheil is the provost at the University of Melbourne.
e-mail: provost@unimelb.edu.au

RESEARCH IMPACT

Income for outcome

Australia and New Zealand are experimenting with ways of assessing the impact of publicly funded research.

BY BRANWEN MORGAN

When it comes to research, governments the world over are asking more questions about whether they are getting value for money — and there is nothing wrong with that, says Peter Gluckman, chief science adviser to the prime minister of New Zealand. "It is what the whole of a nation's science policy process is about: how much to allocate to public sector support; how much to industry sector support; how much to invest in heath versus relative amounts for environment, for instance," he says. "Whether it is done implicitly or explicitly, everyone in that process is thinking about impact."

And whereas large economies have the capacity to invest in a wide range of scientific endeavours, from nanotechnology to cosmology, smaller countries such as Australia and New Zealand do not have that luxury. "The smaller the country, the more limited the choice," says Gluckman. "So when looking at science and innovation systems, you start to be more conscious in the prioritization process; it becomes a much more strategic issue."

Determining the impact of research on wider society has the potential to assist decision-makers within organizations and institutions. But what is troubling people like Gluckman are the definitions. "You have to be really clear about the word: there are many different kinds of impact and perceptions differ," he says. "Governments have to decide what impacts they are looking for." Questions surround what constitutes impact and at what point during or after the research process it should be evaluated. Can something that is subjective and qualitative ever be appropriately measured?

VALUE JUDGEMENT

Gluckman's office in Auckland serves as the administrative headquarters for the science, technology and innovation stream of the Small Advanced Economies Initiative (SAEI) — a network for the discussion of challenges, opportunities and policies that are of particular relevance to small developed nations. For these countries, prioritizing the areas of science and innovation in which they invest is crucial to economic prosperity.

The SAEI has begun to develop an 'impact taxonomy' to help categorize the range of impacts that can arise from research. These include not only the direct economic effects but also intangible factors — for example, a Nobel prizewinner's role in enhancing a country's scientific reputation. Gluckman says that a holistic science system has to consider all the different kinds of impact that matter to people; a taxonomy, he contends, will facilitate that discussion. "What is the 'value' of doing the kind of biosecurity research that makes foot and mouth disease less likely or a country more resilient to an earthquake?" he asks. "This type of research could easily get forgotten if you focus on only one form of impact."

But measuring the breadth of impacts arising from research that often takes place over many years poses tough challenges. Assessors need to identify proxies and intermediate outcomes to gauge the direction a study is taking. That is where Adam Jaffe, director of New Zealand's Motu Economic and Public Policy Research think tank in Wellington, comes in. Jaffe is working on an evaluation framework covering five categories of impact: financial, social, environmental, public policy and

capability. The purpose of this framework is to give decision-makers an idea of the full range of potential outcomes and help them decide which to track — and how.

Many scientific discoveries are serendipitous, and critics suggest that such a goal-oriented approach could lead to less blue-sky research funding. Gluckman disagrees, saying that knowledge advancement is in itself a goal. "To say impact assessment moves you away from basic science is to narrowly use the term 'impact'." Jaffe, an American, uses a baseball analogy to defend impact assessments. "The fact that sometimes you strike out and sometimes you do well doesn't stop us from thinking about who is better on average," he says. "We can look at which models on average generate the greatest outcomes and impacts across a number of different measures." The inevitable randomness in the process, he adds, "doesn't invalidate this approach". Using a framework such as his, which incorporates multiple measures across many dimensions, will minimize the chance of missing important effects.

No single group of measures may indicate both excellence and impact.

Such a framework will probably be attractive to New Zealand's recently launched National Science Challenges (NSCs), which provides designated pots of money to research areas deemed to be of national significance. The NSCs have a funding horizon of ten years — part of the government's move to support science over a longer term than most other funding bodies, which typically give three- to five-year grants. In their proposals, applicants are required to describe their ten-year vision, anticipated outcomes and impact.

Auckland University physicist Shaun Hendy and his team have recently submitted an application for one NSC, called Science for Technological Innovation, which covers work that "enhances the capacity of New Zealand to use physical and engineering sciences for economic growth". Hendy, who is also the director of the Te Pūnaha Matatini Centre for Complex Systems and Networks, hosted by Auckland University, found the requirement to discuss impact hampered his application. "Our proposal covers a very broad range of disciplines and industry stakeholders," he says. "Manufacturing technologies are changing so rapidly that we've struggled to design a research programme that will deliver short-term impact but will also be relevant in a decade."

Research teams that are awarded NSC funding will be required to develop a 'pathway to impact' plan, complete with monitoring and evaluation procedures. Hendy does not believe this sort of assessment is particularly helpful in the long run, because it does not measure the opportunity costs. "To determine

the real benefits of the National Science Challenges, an economist would need to know what we chose not to fund as well as what we did," he explains. "These sorts of exercises are much more about bureaucratic box-ticking than any real attempt to measure the value of science to society".

AUSSIE RULES

The Commonwealth Scientific and Research Organisation (CSIRO), whose headquarters is in Canberra, is the largest of the Australian government's portfolio-funded research agencies. CSIRO is one of the few public research and development entities in Australia — and possibly the world — that formally and transparently plans, monitors and evaluates the impact of its research, according to CSIRO's Mark Johnson. He is project manager of Impact 2020, launched four years ago with the aim of developing a framework to assess the economic, environmental and social impacts of CSIRO's Flagship programmes for use across the organization. For CSIRO, impact is used to influence its "4As": allocation (of resources), advocacy, accountability (to government and other key stakeholders) and analysis (for performance improvement).

CSIRO uses an impact pathway model that describes a project's inputs, activities, outputs, expected outcomes and eventually impact — for example, the adoption of new research protocols that improve productivity (See 'Pathway to impact'). Within the Flagship programme, project leaders can modify these pathway stages depending on changes in resources and goals of either CSIRO or its clients and partners.

CSIRO's research impact planning is a dynamic process; the organization is continually evaluating changes beyond the bounds of a study to see whether the focus of that study is still appropriate. This eye to the wider world also helps with internal engagement. "A lot of people get hung-up on the long-term nature of most impacts, which they see as so far removed from their actions today that they are reticent to engage," says Mark Bazzacco, CSIRO's executive manager of performance and evaluation. He says that "monitoring progress towards impact" as opposed to "monitoring impact" helps scientists to keep an eye on their project's goals and get a greater sense of the role they play in effecting change.

Bazzacco gives an example. Some research impacts are the culmination of decades of work. The Murray-Darling Basin management plan was a large collaborative research project that aimed to measure and model water flows within this vital river basin. In addition to collecting data, the researchers involved also helped policy-makers understand the modelling and the plan's findings to enable creation of new regulations concerning water usage. But once the new policy had

been developed, it was up to state and federal governments to adopt it — and then for many other actors to implement it. It will take another decade at least for the final impacts (environmental, agricultural, social and economic, for example) to be realized; in the meantime, the researchers have moved on to other projects. Monitoring progress towards impact recognizes where the researcher's role is important (for example, in conducting the work and helping communicate the results) and avoids any negative implications of assigning responsibility for behaviours that influence the final impact yet are beyond the researchers' control (in this example, the implementation of the policies).

MAPWORK PROJECT

Over its lifetime, CSIRO has conducted tens of thousands of projects. Johnson is working on ways to show these in an easy to appreciate, visual manner. They have come up with an impact map, shown on page S74, designed to be a conversation starter rather than a precise diagnostic tool. It shows 286 projects — only those that meet a minimum realized or projected financial return criterion and where data are available to evaluate the delivered or intended impact. Impacts can be social, environmental and/or economic. Projects are assigned a primary impact category based on the same 17 socioeconomic objectives used by the Organisation for Economic Co-operation and Development.

The size of the bubbles is based on a four-point scale that allows projects with different types of outcome to be compared with each other. The placement of the bubbles are indicative of when projects have delivered, or will likely deliver, a significant milestone. From 2004, the timescale changes from five- to two-yearly, as there are more data are available for recent projects. Here are some examples, highlighted on the image.

Wireless LAN: CSIRO scientists solved the main problem impeding fast wireless networking of electronic devices — that of reverberation within rooms. The organization applied for several patents and, in 1996, was granted a US patent for wireless local-area network (WLAN, or wifi). This outcome has led to major social and economic impacts and has revolutionized communication. The technology is now used in an estimated five billion devices worldwide. CSIRO has licence agreements, worth more than AUS$430 million (US$400 million) with more than 20 international companies.

Murray-Darling basin plan: The Murray-Darling is the largest river system in Australia. It supplies water to approximately 10% of the population and produces 40% of national agricultural output. Growing demands on its increasingly variable flows have caused widespread concern among communities and industry. CSIRO scientists measured the

available water within the basin and created a model. In 2011 they produced a report on resource planning, management and investment, as well as modelling scenarios for future catchment development, groundwater extraction and climatic conditions out to 2030. The report and models directly informed government investment; the project's impact has been an estimated saving of at least AUS$2.8 billion (US$2.6 billion) through better use of funds and water infrastructure efficiencies.

Tiger prawn and Aquaculture feeds: In 2010, CSIRO announced that its ten-year collaboration with Australian prawn farmers had led to successful selective breeding of Black Tiger prawns with improved growth and survival rates that could be sustainably farmed in salt-water ponds. The estimated economic impact on the industry is AUS$120 million (US$112 million) per annum. Simultaneously, another CSIRO research team developed a new aquafeed ingredient (Novacq), derived from marine microbes, which increases Black Tiger prawn growth rates by 30%. The combined value of the increase in prawn yield is estimated to be AUS$430 million (US$400 million).

The Cooperative Research Centres (CRCs) programme in Canberra, set up in 1990 by the federal government, also looks at the broader impact derived from its applied research. CRCs support multidisciplinary teams, which often include groups from CSIRO, and are carefully managed to deliver impacts, says Tony Peacock, chief executive of the CRC Association, the CRC umbrella advocacy body. All CRCs include participation of end-users from the outset, ensuring that projects are always addressing real-world situations. Any proposal for CRC funding involves completing an impact tool similar to CSIRO's impact pathway approach.

Over the past decade, there have been three independent retrospective analyses of the CRC programme. The methodologies were agreed up front with government officials and use a counterfactual point of view — that is, they compare the impact of the CRCs with a scenario in which each project had not taken place. All three studies showed that CRCs have had a positive impact, including on economic indicators such as GDP. Peacock is open about the intended target audience. "There is no problem convincing the public of

the value of research — they want to hear our stories," he says. "These retrospective studies were directed solely at those who were likely to determine whether the CRC programme continues to get money."

UNIVERSITY CHALLENGE

Australia's higher, or tertiary, education sector, which includes all the country's universities and almost two-thirds of its scientists, does not routinely conduct research impact assessments. These institutions are, however, compelled to monitor research quality through the Excellence in Research for Australia (ERA) initiative (see page S64).

In 2003, the then Liberal government started developing an assessment system — called the Research Quality Framework (RQF) — that looked at the impact of research as well as its quality. But in December 2007, just weeks before the RQF was due to be implemented, an incoming Labor government scrapped it (see page S52). Matt Brown, senior policy analyst in research at the Australian Technology Network of Universities (ATN), in Adelaide, believes that an opportunity to provide a useful decision-making tool was lost. "We've seen

Delivering positive impact

CSIRO's National Research Flagships are taking on the biggest challenges and opportunities in manufacturing, minerals, energy, digital services, water, agriculture, food and nutrition, oceans and atmosphere, and biosecurity. This map represent the key impacts CSIRO has delivered and intends to deliver. Some impacts will not happen, new discoveries will make others possible, and Australia will ask CSIRO to respond to new challenges. CSIRO does not deliver impact alone – we work with more than 2000 Industry, Government and Research partners each year to create a better future for Australia and humanity.

Organizations worldwide wrestle with the issue of how to show and compare the impact of their projects. Australian national science agency CSIRO is working on an impact map. Each of the 286 bubbles represents one project with one primary category. The four sizes of circle relate to size of impact (economic, social or environmental), centred on when projects met, or will meet, a significant milestone. Larger version: go.nature.com/bduk6p. Website: csiro.au/impact.

car manufacturing suddenly disappear from Australia and now the country is asking what its future industries will be," he says. "Research impact assessment would have helped guide today's decisions."

Another attempt to push the research agenda came in 2012, when 12 Australian universities took part in the Excellence in Innovation for Australia (EIA) trial. The EIA demonstrated how a case-study approach could be used to systematically assess the impact of university research across a wide range of disciplines and areas. "The time was right to have another go at putting impact on the agenda," says Brown. "And when trying to convince policy-makers, you need proof by demonstration." In the EIA trial report, the authors note that they met their objective to "measure the innovation dividend of research generated by Australian universities" and advanced the methodology to do so.

Australia's chief scientist, Ian Chubb, seems undecided as to the benefits of an impact assessment system for the tertiary sector and has concerns around the retrospective case-study approach, as recently implemented in the United Kingdom's new Research Excellence Framework (which, ironically, was influenced by the aborted RQF). This approach is also highly selective, with universities submitting only those studies that demonstrate the best results. Chubb echoes Gluckman in the need to articulate the goal of any assessment system. "If the purpose of measuring impact is to show politicians just how much scientific research underpins sectors of the economy, then we should own up to it," he says. "But if impact evaluation is to be tied to funding, then you have to be careful that you're not just being fashionable. You're not going to give a university funding because 15 years ago someone had an idea that turned out to be profitable."

In an effort to ensure better coordination of activity and investment, without succumbing to the whims of fashion, the Australian Research Committee (ARCom), based at the Department of Industry in Canberra and chaired by Chubb, put together a national research investment plan to guide government. The plan was released in November 2012. As part of its plan, ARCom, which includes the Canberra-based chief executives of the Australian Research Council (ARC), the National Health and Medical Research Council (NHMRC) and CSIRO, recommended the development of a university-focused impact assessment mechanism as a companion to the ERA. ARCom released a discussion paper in June 2013 and the submissions window was closed in August — just in time for another

"If the purpose of measuring impact is to tell politicians the value of research, let's own up to it."

PATHWAY TO IMPACT

Each of CSIRO's Flagship projects is guided by this framework, which gives project leaders a way to think about their work so they can plan and monitor for impact.
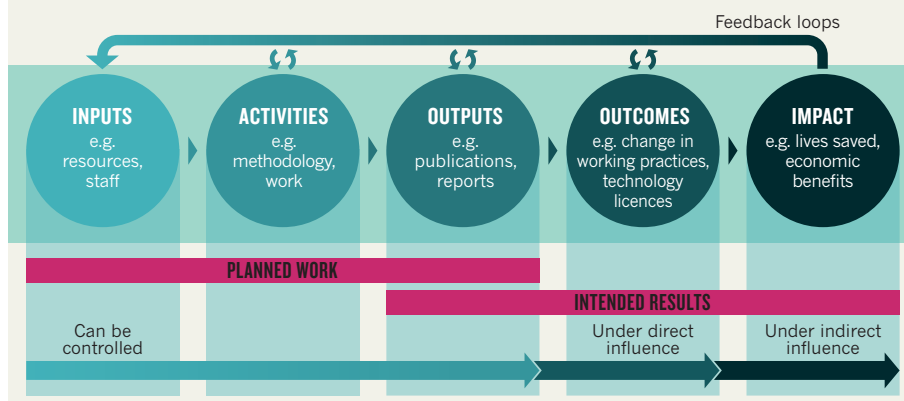
change in government. The incumbent Liberal party has not endorsed the plan.

In the absence of government leadership, one Australian university has decided to go it alone. "We are in a different political environment now," says Warren Payne, pro vice-chancellor in research and research training at Victoria University, Melbourne, "and we need to be ready for a measure of impact that might come on top of the current ERA framework."

Individual research programmes at Victoria University are assessed for retrospective and prospective impact; economic, social and environmental outcomes are combined with qualitative surveys that gather the views of people who are socially or financially invested in the project on potential future impacts. The university is currently trialling the system and has brought in an independent assessor to evaluate progress. So far, 11 projects from Victoria University's social science and science technology departments have been successfully assessed via a rating scale that considers the significance and reach of each of the impact claims.

Payne sees Victoria University's impact tool as a useful way to guide its research decisions. But he says that perhaps the greatest attribute of the system is to focus the minds of researchers as to what expectations they are raising and whether they are delivering on them. And, as an added benefit, they are more motivated because they understand how their work feeds future impacts.

GROUP EFFORT
One of the strongest arguments for assessing impact is the inclusion of research that falls outside the traditional criteria for academic excellence. "Work that previously might have been seen as highly applied — almost in a pejorative sense — might then be recognized as being important," says Tim Wess, executive dean of science at Charles Sturt University in Wagga Wagga. Wess gives the example of research into changing a nursing procedure that reduces post-operative mortality, compared

with fundamental research in particle physics. Assessing both types of research by their excellence and impact "would level the playing field", he adds. University promotion committees and grant application reviewers could also take this information into consideration.

Unfortunately, no single group of measures may be able to indicate both the excellence and the impact of research. Publishing a paper, even an 'excellent' one, may not have an impact outside academia without additional effort — often undertaken by others — to translate that knowledge into practice. This raises the issue of timeframe. The wider benefits of research might not appear for years or decades, by which point many individuals and organizations may have contributed. How can an impact measure tease out separate contributions?

Indeed, coming up with a suitable methodology is a big sticking point. Payne has discussed Victoria University's approach with the ARC, which administers the ERA assessment exercise and is the larger of Australia's two research-funding councils. "ARC feels that although our method can probably be scaled quite nicely within the university, it isn't sure it can be done systemically," he says. The other funding body, the NHMRC, requires some mention of impact in its grant proposals, but the guidelines are not explicit and vary by scheme. "Consensual views as to what can be claimed as impact and what evidence should be provided would be very helpful," says NHMRC chief executive Warwick Anderson.

So far, it is a discussion that has been disparate — and one that, given its membership, ARCom seems well placed to continue. "I'm not saying it shouldn't be done — or that a whole bunch of smart people can't find a way," says Chubb. "We just can't afford to be simplistic about it." ∎

Branwen Morgan *is a freelance science writer based in Sydney, Australia.*

PERSPECTIVE

BRETT D'ELTON

Powering up citations

Changing the way we measure and reward research could enrich academia and improve outcomes for society, says **Alan Finkel**.

Australian science suffers from a fundamental misalignment. Publicly funded researchers at universities face considerable pressure to generate academic papers. Taxpayers, however, would prefer to see more significant commercial and social benefits from their research investment.

At the national level, the Excellence in Research for Australia (ERA) process, through which the federal government evaluates universities, is driven by assessment of the quality of research publications. In the science, technology, engineering and mathematics (STEM) disciplines this assessment is based on citations — a score measuring a published paper's influence by the number of other papers in which it is cited. In the humanities and social sciences (HASS), the assessment is based on academic peers reading selected papers to determine quality. The ERA does credit other accomplishments[1], including fellowships of learned academies, patents and registered designs, plant-breeders' rights and research commercialization income, but there is little evidence that assessments have given much weight to such achievements.

Because faculty members, departments and universities want to be judged as being world class or better in the ERA assessment, they pursue research for academic publications that are likely to be well cited — almost to the exclusion of other activities. And since evaluation programmes such as the ERA affect funding and student demand, they drive academic behaviour.

Individual researchers realize that the path to promotion is paved with academic papers, and so rarely spend time working on anything else. For example, academics have little incentive to join in industry programmes such as the successful Cooperative Research Centres (CRCs) if participation limits their ability to publish extensively. Universities and research institutes considering the appointment of an academic who has spent years in industry often worry that the applicant's grant-winning ability might be compromised by his or her time away from academia. Similar issues dog researchers within government research institutes such as the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

These problems thwart engagement with industry, thus depriving researchers of useful commercial skills. In Australia, more than twice as many PhDs are employed in universities than in industry, whereas in Germany the ratio is the other way round[2]. As a result, Australia was rated last in the recent ranking from the Organisation for Economic Co-operation and Development (OECD) of university–industry collaborations[3]. Without such collaborations, industry does not fully benefit from the fundamental research undertaken in academic settings, and academic researchers are not as aware as they should be about market and societal needs and trends.

One way to encourage academics to collaborate more with industry would be to award 'citation equivalents' to various activities that advance the practical impact of science through means other

TAXPAYERS WANT TO SEE **COMMERCIAL** AND SOCIAL **BENEFITS** FROM THEIR RESEARCH **INVESTMENT.**

than peer-reviewed publication of academic papers. In STEM disciplines, citation equivalents could be calculated for issued patents, commercial contracts and licence fees. More broadly, citation equivalents could be awarded for activities including writing books, opinion pieces and government submissions, PhD student supervision, and development of new approaches to teaching practices or novel training courses.

Citation equivalents earned could be counted in the same way as normal citations, even contributing to higher order measures such as the H-index (a measure of a researcher's impact and productivity) or institutional-level evaluations such as the ERA. Each contributing activity would count as equivalent to a paper with an agreed number of citations. For example, an Australian patent might be rated as equivalent to a paper with a small number of citations, say five. On the other hand, a triadic patent (which covers the US, Europe and Japan) might be rated as equivalent to a paper with 50 citations; patents taken up and used would be rated more highly than ones that lie dormant.

A system such as this, aiming to provide impact measures for individual effort, would be cheaper and faster than the labour-intensive methods that are needed to gauge institution-level impact. Case studies and expert evaluation panels take a long-term view, in some cases considering outcomes a decade or more after publication. The proposed citation equivalents, by contrast, would measure near-term achievements as soon as the impact activity is definitive, for example, the issue of a patent.

Because traditional citations are global in extent, citation equivalents could be considered for adoption not only in Australia, but worldwide. To be successful, citation equivalents would have to be embraced by research institutes, universities, national granting agencies and, ideally, international evaluation programmes and databases such as Scopus, Google Scholar and Web of Science.

Citation equivalents could be tested on a small scale and rolled out as experience is gained. With a little funding and some determination we could broaden the existing publications-focused metrics to achieve a better balance — acknowledging the best basic research while also promoting the STEM research that delivers the greatest impact for the society that is paying for it. ∎

Alan Finkel *is president of the Australian Academy of Technological Sciences and Engineering and chancellor of Monash University.*
e-mail: alan@finkel.net

1. *Excellence in Research for Australia 2012, National Report* (Australian Research Council/Commonwealth of Australia, 2012); available at go.nature.com/fcunry
2. Pettigrew, A. G. *Australia's Position in the World of Science, Technology & Innovation.* Occasional Paper Series, Issue 2 (Australian Government, 2012); available at go.nature.com/x4yp7g
3. *OECD Science, Technology and Industry Scoreboard 2011* (OECD, 2011); available at go.nature.com/n914ae.

PERSPECTIVE

If not funding then teaching

The lack of financial reward from Australia's national system of research assessment is obscuring the real issue, says **Brian Schmidt**.

When he announced the formation of the Excellence in Research for Australia (ERA) initiative in 2008, Kim Carr, then minister for innovation, industry, science and research, said that it would provide a "transparent, workable system to assess the quality of home-grown research". Carr strongly hinted that future funding decisions affecting higher education institutions would be informed by outcomes of the ERA, which would collect data and rate the quality of their research output.

For a comparable model, Australian universities looked to the equivalent system in the United Kingdom, the Research Assessment Exercise (RAE), which allocated significant fractions of the available research money to universities on the basis of their RAE scores. The implication for Australia was that a poor score would lead to financial disadvantage.

Six years later the sector is gearing up for the third instalment of the assessment process. But during this time, the money that was supposed to be tied to ERA outcomes has all but vanished. The incentive structure to award the money is in place: on a scale of one to five the two lowest ratings attract nothing, whereas the top rating (five, or 'well-above world-standard') earns seven times that of the middle rating (three, or simply 'world-standard'). But the total amount of money available is trivial compared to the overall budgets of the participating universities. In fact, ERA financial reward accounts for only 1.2% of Australia's investment in higher education research and development (HERD), or just over AUS$116 million (US$109 million).

Given that the government has spent AUS$43.5 million on the ERA, and universities themselves have outlaid substantial sums to undertake the ERA evaluations since 2008, one might question the value of this exercise that awards so little money.

NON-FINANCIAL INCENTIVES

Nonetheless, it is clear that ERA has helped influence the AUS$9.6 billion invested annually in HERD. By focusing its assessment on research quality, rather than quantity, the ERA has helped elevate the research at many of Australia's universities (see page S67). The sector is now strategizing about research quality — and these plans are manifested in new initiatives across various universities. There is evidence that excellence is being recognized and rewarded as one of a series of outcomes.

ERA's impact has gone beyond universities and helped to measure Australian capabilities against benchmarks across the breadth of the HERD sector. This is a useful exercise that should help Australia invest more strategically in research in the future — a necessity thrown into sharp relief by the 2014 budget, which introduced big changes for the higher education sector. Among other items, the budget removed the cap for university tuition fees bringing potentially profound implications for the higher education sector. In this instance the ERA process can provide a benchmark to gauge the effect on research quality. The results, positive or negative, can be used to inform policy decisions around the impact of the decision to deregulate fees as well as other reforms proposed in the future.

But how much influence can the ERA continue to wield once the HERD sector realizes that the total pool of money on the table is tiny? Some commentators have said that, without significant funding flowing from ERA rankings, the programme is not worthwhile — but this is not where the problem lies with funding for HERD.

A far more sensible system is one that contributes towards the full cost of research as part of the granting process, as happens in the United Kingdom, United States and Canada, for example. An assessment system like ERA would then give additional strategic money to help institutions do even better research, at a level in line with current funding. Unfortunately, Australian grants provide nowhere near the full cost of research; significant cross-subsidization is required from student fees. This undesirable method of research funding is unfair to students who believe they are paying for their education but are in fact paying for the country's research.

> A FAR MORE **SENSIBLE** SYSTEM IS ONE THAT PROVIDES THE **FULL COST** OF RESEARCH IN THE GRANTING PROCESS.

As a fraction of GDP, Australia spends more on research within higher education than most of the countries in the Organisation for Economic Co-operation and Development (OECD), but its overall rate of R&D investment is well below the OECD average. It is therefore important that Australia maximizes its returns from research within the higher education sector. ERA has successfully emphasized research quality, but this is against a dearth of assessment on how our universities interact with industry (see page S77).

Given this set of incentives, it is perhaps unsurprising that although our universities' research outputs are ranked eleventh in the world, Australia was ranked last for business collaboration with higher education and public research agencies within the OECD. Australia needs to invest more in R&D, but without a strategic plan to achieve educational and business outcomes in tandem with excellence in research as captured by ERA, our country will not fully benefit from its investment.

The next round of ERA evaluations is scheduled for 2015. And although the ERA has been worthwhile, it is unclear how much is to be gained by undertaking this formidable exercise again so soon. Not much has changed in the past three years in the Australian HERD sector, so this triennial exercise — as it stands — seems too frequent. On the other hand, if ERA can spur the government to strategically plan its research agenda, then supporters and naysayers alike would rejoice in being assessed as often as is deemed necessary. ■

Brian Schmidt is professor of astronomy at the Research School of Astronomy & Astrophysics, Australian National University, Weston Creek, Australian Capital Territory.
e-mail: brian@mso.anu.edu.au

Q&A Jane Harding
Individual approach

Jane Harding is deputy vice-chancellor for research and professor of neonatology at the University of Auckland, which is New Zealand's most well-funded university under the Performance-Based Research Fund. She discusses the country's approach to assessing science and measuring impact, and describes why she prefers a model that grades the individual not the research group.

What do researchers in New Zealand think about the Performance-Based Research Fund (PBRF)?

Most researchers see the PBRF as an inevitable chore. Still, although they find it time-consuming and distracting to prepare a portfolio, they see some advantages in having something that provides an external validation of research performance on a regular basis — in our case, every six years.

Although many researchers believe that the PBRF is a rather imperfect measure of research quality, it is, by and large, a useful thing. It would be difficult to find an alternative way to distribute the money to the most research-intensive groups. And there is evidence that the PBRF has improved the quality of research.

What is your involvement with the PBRF?

I am responsible for running the whole process within the University of Auckland. The early stages are mostly about preparation and education — making sure that everybody knows what they have to do. Then we get into the phase of assembling individual portfolios. We run an internal review of the draft portfolios before they get polished for final submission, and when the assessment is completed, we manage the results and make sure that they get to the right people.

It is a very big, complicated and continuous process involving a lot of human resources work: we have close to 2,000 portfolios to submit by a specific date. At its peak it can occupy about half of my time for two to three weeks.

This seems like a significant investment for the university.

Yes — but it is also very important to us. The PBRF contributes about 8% of our total budget, which is a significant chunk of funding. Universities in New Zealand did a rough approximation after the last PBRF round on the costs of the process to universities; our best estimate was less than 3% of the total PBRF income over the six-year period, which is not a huge overhead.

What is the grading experience like for researchers?

From an institutional point of view, the distribution of individual grades — A, B, C or R (for research inactive) — has little overall effect on us in terms of dollars. If one of our researchers is awarded a B instead of an A, that is usually balanced somewhere else in the institution by somebody getting an A instead of a B. Such variations don't make much of a difference to the profile of the institution, but they make a huge difference to the individuals. Getting an assessment of a B when you thought you might have been eligible for an A is a huge disappointment. People will inevitably interpret the grades as defining something about themselves; you can't stop them from taking it personally.

Have any categories of researchers been disadvantaged by the assessment?

The assessment is experience-dependent, which makes it difficult for a junior researcher to get an A grade. A brand-new postdoctoral researcher is not going to have a strong research portfolio. The PBRF has a category for new and emerging researchers, in which the threshold for getting a C is much lower, and this does mitigate some of the disparity.

You could even argue that researchers with less experience benefit more from the PBRF because there is assistance at an institutional level: the PBRF provides institutions with funds to specifically support supervision for research degrees. In our university, the PBRF has contributed to an increase in support for less-experienced researchers.

Does the PBRF have a bias against specific research fields or types of output?
That is what everybody worries about, but I don't think it is a reality. You have to trust that the reviewers can assess the different kinds of research in an appropriate way. A few concerns have been raised about the criteria for 'world-class' research, which could disadvantage disciplines focused on indigenous research in local communities. But any bias would result from a misconceived equating of world-class research with international research. You can do world-class research on New Zealand topics.

There is also a minor concern that some types of research, for example those involving commercially relevant work with private companies, might be discouraged because it doesn't necessarily result in a report that can go into a portfolio.

But overall, I don't think that the PBRF has changed the nature of scientific inquiry. It does create pressure to produce outputs, which means that people who would like to sit and spend 15 years writing a book are not going to do well in the PBRF, but they won't do well in any environment that is focused on research quality.

Overall, more than 10% of funding to universities in New Zealand comes from the PBRF. This is much lower than the 25% allocated by a similar programme in the United Kingdom, but higher than the 2% allocated in Norway. Do you think that enough money is distributed through the PBRF?
The amount distributed through the PBRF — NZ$262.5 million (US$224.2 million) in 2013 — is enough to provide a significant incentive but not enough to cover the costs of the research that it is designed to support or to ensure the highest quality research. However, the international comparison comes down to how other components of the system are funded. Universities in New Zealand are seriously underfunded by any measure. We have one of the lowest funding rates per student compared to other OECD [Organisation for Economic Co-operation and Development] countries. And our expenditure on research and development as a proportion of GDP is about half the OECD average.

As a result of this shortfall, researchers in New Zealand spend a substantial amount of time seeking funding. We also face difficulties in recruiting and retaining good researchers. Even though a world-leading professor may

The University of Auckland was allocated more than NZ$72 million from the PBRF in 2010.

be interested in coming to New Zealand and accepts our salary levels, he or she is often discouraged by the lack of research funding and so might choose not to come here. That is a major disadvantage. More funding through the PBRF would help to support more research, but would not make up for the serious underfunding across the sector.

How does New Zealand's approach compare with other peer-review-based models such as in the United Kingdom?
The difference in New Zealand is that the assessment is done at the level of the individual as opposed to the research group. Arguably, an individual-based system could lead to selfish behaviour because there is no direct incentive to work collaboratively and to support a team or more junior researchers.

In New Zealand, this is counterbalanced by requiring that portfolios include not only research outputs, but also evidence for the section called Contribution to the Research Environment — a category that covers activities such as engaging in peer review activities, leading collaborative groups, supervising postgraduate students and mentoring early career researchers. It would be very difficult to apply the individual portfolio model to a larger system because of the scale of the assessment.

I prefer the individual model. The UK process requires gathering all of the individual material and then assembling that into an aggregate submission, so it seems to be a lot of additional work for not a lot of additional gain. My colleagues in the UK talk about their universities employing people full-time just to write the submissions.

And how does it compare with indicator-based models like those in Denmark and Australia?
Any peer-review process is vulnerable to the vagaries of individuals. Assessments can vary based on who is on the panel, how well they know the subject, their own personal prejudices, as well as many other unquantifiable factors such as how well the portfolios have

been written. The system is also expensive, because each individual portfolio in New Zealand needs to be prepared and assessed.

A metrics-based system is much cheaper and simpler to run and would come out with almost the same outcomes if one was simply talking about allocation of the money and alignment of institutional goals to research quality objectives. But New Zealand's approach of submitting individual portfolios brings the incentives back to each individual staff member in a much more direct way than does submission of metrics at the institutional level. There is also the issue that the metrics themselves, rather than research quality, can become the target — and the lack of peer review means that different disciplines might be differentially treated.

> *"Researchers with less experience benefit because the PBRF provides funds to support supervision."*

Should the PBRF include measurements of impact, similar to the United Kingdom's new Research Excellence Framework?
It is challenging to get a single system to measure two different things. If you consider research quality and research impact to be separate things, then you need separate processes. It is difficult to measure impact — and expensive in terms of the effort required to assemble the evidence.

Introducing new measurements would be useful if they created incentives for academics to increase the impact of their research, but impact is so closely related to research quality that I am unconvinced as yet that a separate assessment is worth the enormous cost. We will learn from the UK's attempt to assess impact on a national scale. ∎

Interview by **Smriti Mallapaty***, assistant editor for Macmillan Science Communication in Tokyo.*

nature

http://www.nature.com/nature/outlook/assessingscience

Photo: Aussie Bacteria - Early stage of biofilm development by Pseudomonas aeruginosa showing a single microcolony (L. Turnbull, C. Whitchurch, ithree)

THE VIEW FROM INSIDE A MICROSCOPE

Monash University has built the microscope of the 21st century, a cinematic virtual reality environment that gives researchers a new spatial awareness of their subjects.

Dubbed CAVE2™, it comprises a curved room with 80 high-definition LCD screens projecting 3D images that allow a viewer, when wearing special "tracked" control glasses, to not only see their subject but also to manipulate and walk through and around it.

"This opens up new possibilities for insight and discoveries across a range of sciences," says Professor Paul Bonnington, director of the Monash eResearch Centre (MeRC), who has overseen the development of CAVE2 at Monash University's Clayton campus in Melbourne.

For example, the brain's white matter viewed up close in CAVE2 can provide biomedical researchers with new insights into disease—especially when they are able to compare diseased and healthy samples.

Senior research fellow and CAVE2 platform manager Dr David Barnes explains: "If you looked at this on a desktop display it would basically look like a bundle of wool. You don't appreciate the space and gaps between the circuitry and its 3D structure, but CAVE2 lets researchers see the actual structural differences."

The technology behind CAVE2 was developed by the Electronic Visualization Laboratory at the University of Illinois at Chicago in the US, and their experts helped install it in Melbourne, where it's been operating since November 2013.

But it's the additional work by Monash University, integrating CAVE2 with other

CAVE2 at Monash University displaying brain white matter data courtesy of UIC. Photo: Paul Jones

advanced imaging facilities, that has made it into the modern microscope.

MORE THAN THE VIEWFINDER

A microscope viewfinder makes a good analogy for CAVE2, according to Professor Bonnington. At the centre of scientific discovery for hundreds of years, a microscope has three key components: at the bottom, a light source to illuminate a sample; in the middle, the focusing dials; and at the top, an eyepiece for viewing.

"Five years ago we set out saying, 'scientific discovery is still going to depend on this concept, but we need a modern equivalent'," he says.

In his 21st century version, the "light source" is the imaging technology that provides the sample. It could be instruments like the Australian Synchrotron, a magnetic resonance imaging scanner or a next-generation DNA sequencer.

The "focusing" components of the traditional microscope are replaced by computational tools that transform and filter the data sample to extract features. These are provided by a purpose-built interactive supercomputer installed at Monash University called the Multi-modal Australian Sciences Imaging and Visualisation Environment, also known as MASSIVE.

The Centre also developed their own data transport, management and storage software called MyTardis, which takes the huge amount of data from the bottom instrument layer of the "microscope" and places it into MASSIVE.

Finally, the modern-day "viewfinder" is a powerful viewing lens—CAVE2 itself—which can reveal features, details and a perspective never before possible.

SEEING WHAT WE'RE MADE OF

These possibilities include building further on the Human Genome Project to

examine the finer details of the human proteome, the complete set of proteins expressed by our genes.

"Mapping the human genome was really just prep for where the research is going, and that is to understand, not just the building blocks, but how they function," explains Professor Bonnington.

The Human Proteome Project is providing this understanding by mapping proteins to shed light on protein function, and to advance the treatment of disease.

A protein's function is often determined by its shape, and how this shape can change. And there is no better place to analyse protein shape than in CAVE2, where, for example, electron microscopy data converted and modelled through MASSIVE can be studied up close in a super-sized format in three dimensions.

Similarly, CAVE2 could be used to fast-track drug design. "When you know the structure of a target molecule you could come in here with a bunch of drug candidates and literally carry them over to the molecule and see if they fit," Dr Barnes says.

He says the human brain is still the best pattern-recognition tool we have, which is why viewing data in high-resolution, virtual-reality detail in CAVE2 is such an extraordinary opportunity for discovery.

TRANSCENDING TIME AND SPACE
Medicine is not the only field to benefit from CAVE2, and the instrument can display more than just the microscopic world. Planetary scientists are able to step inside a panorama of

Mars that has been reconstructed from images originally taken by the NASA rover Curiosity.

As with the biomedical images, seeing the planet's features at scale and looking real enough to touch gives new insight for researchers. "You can deduce a lot more from this perspective," Dr Barnes says.

Dr Barnes also says CAVE2 can assist research and industry across many other fields. Archaeologists, for example, can study fragile historical ruins in detail without disturbing them. And engineers can safely observe how a severe storm would batter infrastructure.

"We have had more than 1500 visitors from research and industry," Dr Barnes says. "People walk away feeling inspired to imagine new uses for the cave in their own fields."

VIRTUAL ENGINEERING
One imaginative new use was modelling crowd flow and bottlenecks at the redeveloped Second Avenue Subway precinct at Fulton Street, the site of the former World Trade Center towers in New York.

"You can actually see people moving around you in their avatar form, and you get a greater understanding of how you can improve the physical operations of a building when people are there," says Peter Bowtell, buildings practice leader of global design and engineering firm Arup Australia. "It is all about being able to visualise big data sets in a meaningful way."

Arup is also using the 3D format of CAVE2 to give clients and practitioners the ability to walk through and around

a virtual finished building and experience what it will look, feel and even sound like.

More than just a "wow factor", this lets engineers test the impact of design choices on noise level and identify conflicts with plumbing or electricity conduits, avoiding costly revisions later. Mr Bowtell says "virtual construction" like this is driving anticipated savings of 15 to 20 per cent in the US and the UK.

"You can see new ways in which CAVE2 could really be incredibly valuable for proving building and project concepts before and during their production cycle," he says.

Professor Bonnington says the potential for application and collaboration with research and industry—and as a teaching tool—is boundless.

"The technology itself inspires people," he says. "They see technology used in a way they have never seen before, and they can begin to imagine what the future could be like."
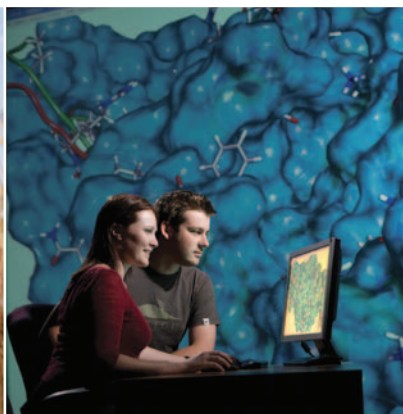
For further information on CAVE2 please email Professor Paul Bonnington: **paul.bonnington@monash.edu**

To be kept informed of our latest research developments, subscribe to Monash Magazine online. Visit: **www.monash.edu/monashmag**

Monash University
www.monash.edu

THE UNIVERSITY
of ADELAIDE

A CULTURE OF INDEPENDENT THINKERS

Some universities might be content with fostering Prime Ministers and Rhodes Scholars, Nobel Prize winners and astronauts. But the University of Adelaide has its sights set even higher—to be a world-leading research institution, meeting the challenges that will reshape lives and ecosystems.

As one of Australia's most research-intensive institutions, the University has a proud history that goes back to the youngest ever Nobel laureate, Sir William Lawrence Bragg, who attended the University of Adelaide in the early 1900s and who, when aged 25, shared the Nobel Prize in Physics with his father in 1915.

"We have a tradition of exceptional research and a research culture generated by a long history of independent thinkers," says Pro Vice-Chancellor (Research Strategy) Professor Robert Saint. "It is a broad-based institution and we do a wide range of research, from basic through to the applied."

Consistently ranked in the top 1% of the world's universities, Adelaide has 60 research areas rated at or above world standard. It also ranks among Australia's top three universities for contract research and commercialization activity, demonstrating strong connections to external partners.

A HEALTHIER START TO LIFE

At the Robinson Research Institute, researchers are taking a 'bench to bedside' approach to providing children with the healthiest possible start to life. The Institute's transdisciplinary focus is delivering results, particularly in the vexed area of obesity prevention.

"We have a very strong interest in tackling intergenerational transmission of obesity; preventing the obesity epidemic by taking the new strategy of stopping it at its root rather than trying to reverse the situation when people are already overweight," says the Institute's Director, Professor Sarah Robertson.

Bringing together laboratory studies, population-based cohort studies and clinical trials, the Institute's researchers have found that obesity is programmed into a child as early as conception, with both parents contributing equally to a child's metabolic destiny, through their genes and the conditions of the reproductive environment.

"It means that parents have a major effect on the life course of their offspring right from the time of conception, when most people don't even realize they're getting pregnant," Professor Robertson says.

The discoveries offer potent new targets for obesity prevention. Already, principles learned in animal studies are being borne out in large-scale human clinical studies and trials of interventions aimed at breaking this pathway to obesity. Results of some trials have already been published in high-impact journals, while other studies are ongoing.

NEW HORIZONS IN SENSING

The 1966 science fiction film *Fantastic Voyage* envisaged miniaturization technology that could send a medical team traveling around inside the body.

While it might be impossible to send tiny people into the body, scientists at the University of Adelaide's Institute for Photonics and Advanced Sensing are developing sensors at the nanoscale to conduct real-time tests on single cells within living organisms.

In collaboration with Macquarie University, the institute's researchers have developed optical fibers laced with nanocrystals, which are small enough to interact with targets at a cellular level. The aim is to develop a miniaturized laboratory that can operate directly with cells *in vivo*.

The next step will be to use this approach to make meaningful measurements in the brain, the cardiovascular system and even developing embryos, says Institute Director Professor Tanya Monro.

"Any assay that would normally be carried out in the lab could be done at the level of a single cell, without taking it out of the body," Professor Monro says.

PROTECTING A FRAGILE ENVIRONMENT

The global timber industry is worth an estimated $180 billion each year. Yet approximately one-third is illegal. An effective strategy for cracking down on this illegal trade, and thereby reducing associated problems of tropical deforestation and greenhouse-gas emissions, is to empower consumer choice by identifying the source of a piece of timber.

Researchers at the Australian Centre for Evolutionary Biology and Biodiversity have developed a DNA-based tracking method that uses genetic analysis to pinpoint the species and geographic origin of a piece of timber with remarkable accuracy. This technique can reveal whether a piece of merbau timber (*Intsia palembanica*), for example, has been harvested from a certified plantation or whether it has been taken illegally from an uncertified source, even within the same country.

As large-scale DNA screening becomes easier and cheaper, and can be conducted at any stage along the supply chain, a reference database of DNA samples from timber is being compiled to ensure easier and more accurate matching of timber products to their source.

This DNA-based tracking system—the only one of its kind available in the world—has now been commercialized in partnership with the Thünen Institute in Germany and Double Helix Tracking Technologies in Singapore, and it is proving invaluable for verification of sustainably harvested timber products.

"There are a range of certification methods being used, but certification papers can be falsified," says Centre Director Professor Andrew Lowe. "The role of DNA analysis is to verify those certification claims."

CHANGING HUMAN POPULATIONS

We live in an era of unprecedented demographic changes, with migration altering the face of nations, and changing climates triggering the relocation of populations.

The University of Adelaide's Professor Graeme Hugo is one of the world's pre-eminent demographers. As director of the Australian Population and Migration Research Centre, he is at the forefront of exploration into how our global population is changing, evolving and moving, and the economic and social impacts of these shifts.

The University's long established linkages with Asia have helped make it the largest research and training Centre in migration research in the Asia-Pacific region, with more than 50 scholars from countries in the region gaining PhDs in migration and population issues, and more than 20 currently studying.

"The University has become an important regional centre for its holdings of population and migration data from the region, and its staff have high levels of skill in analyzing them," Professor Hugo says.

"THESE ACHIEVEMENTS EXEMPLIFY THE CAPACITY OF THE UNIVERSITY TO TURN ITS HIGH-QUALITY FUNDAMENTAL RESEARCH INTO OUTCOMES THAT CREATE A BETTER FUTURE"

As well as examining demographic changes at the national and regional levels, researchers are also working with global organizations such as the Nansen Initiative to study the impact of climate change on migration.

"Climate change is going to create massive disruptions and displacements of people," Professor Hugo says. "But in fact there is a much more complex relationship in which the effects of environment and climate change are going to be part of a wider pattern of economic change."

GROWING NEW FUELS

With the twin challenges of climate change and peak oil looming, the race to find viable, sustainable alternatives to fossil fuels has never been more urgent.

In response to this need, the University of Adelaide's School of Chemical Engineering partnered with Murdoch University to establish Muradel Pty Ltd, which has developed Australia's first commercial-scale demonstration plant for the production of green crude from microalgae. It's another illustration of how the University encourages connections that reach from basic research to commercialization.

"The challenges with the production of biofuels from biomass are sustainability and high productivity," says Associate Professor David Lewis, one of the project's lead researchers and CEO of Muradel Pty Ltd.

Using a strain of microalgae isolated by Murdoch University, which grows rapidly in saline water, the team has been able to upscale an initial pilot plant, while the University of Adelaide's chemical engineers have developed the vital technologies that allow concentration and extraction of the dilute algae into a commodity product.

A demonstration plant has been built and commissioned in Whyalla, South Australia, with plans to build a larger thousand-hectare plant that will be able to produce around 450,000 barrels of green crude per annum—enough to supply a medium-sized industry such as mining, trucking or a regional airline.

This project, and the many others like it, are proof of the University of Adelaide's capabilities, says Professor Robert Saint.

"These achievements exemplify the capacity of the University to turn its high-quality fundamental research into outcomes that create a better future for local communities and the wider world." ▬

THE UNIVERSITY of ADELAIDE

The University of Adelaide
www.adelaide.edu.au

PURSUING PLASTIC POLLUTION AT THE ATOMIC LEVEL

Plastic pollution is finding its way into the tissues of marine wildlife. Although many environmental risks of plastic have long been known, analysis by the Australian Nuclear Science and Technology Organisation (ANSTO) has detected chemicals from plastic pollution in the feathers of seabirds. However, this research is also giving clues to how to manage the problem.

The work is a collaboration between Richard Banati, a biomedical scientist at ANSTO LifeSciences, and Monash University conservation biologist Jennifer Lavers.

They analysed the elemental composition of plastic items collected from the stomachs of flesh-footed shearwaters. These were then compared with the atomic elements in the feathers from birds of the same species—some that had eaten plastics and some that had not.

The plastic and feather samples were analysed first at the Australian Synchrotron light source in Melbourne and then at the country's only operating nuclear reactor, the ANSTO Open-Pool Australian Lightwater (OPAL) research reactor in Sydney. There they were subjected to neutron activation, in which atoms are bombarded with neutrons to make them slightly radioactive.

Each element is then identified by its distinctive gamma-ray spectrum, giving measurements of composition so accurate they've been compared with determining the vintage of a single glass of red wine spilled into Sydney Harbour.

This produced some interesting findings, such as a regular distribution pattern of elements in each feather— much like the rings of trees. It also confirmed the researchers' initial suspicions: trace elements from plastic were found in the feathers of birds that had eaten it.

The health effects of these elements aren't well understood, but their presence in tissues shows the complexity of the pollution problem. Contrary to expectations, plastics that degrade in the environment are not necessarily safer for wildlife, as the increased surface area as they break down can exacerbate the release of toxins like cadmium and mercury.

"A traditional approach to environmental management has been 'the solution to pollution is dilution,'" says Richard Banati. "However, we are finding that mass plastic consumption, together with increased degradability of plastics, may actually lead to a steady increase of hazardous contaminants in the environment which would be difficult to reverse."

However, the researchers stress that their aim is not to demonise plastic, but rather to better understand it at an atomic level.

Richard Banati, who has used radioactive elements in pharmaceutical research to track a drug's progress through the body, points to the intriguing possibility that manufacturers, too, could add an isotopic signature of non-radioactive trace elements to plastics to trace their lifecycle.

A combination of non-abundant elements together with a small amount of precious metal like gold—as little as 10 milligrams per tonne—could be detected by ANSTO's instruments and determine the litter's original source.

"Plastics will always have a place in our world," says Richard. "But we need to be mindful that a seemingly 'single-use' throw away item will change form many times and stay in the system at the atomic level for eternity."

ANSTO's Institute for Environmental Research already uses the ability to distinguish naturally occurring isotopes to follow the movement of groundwater and rainwater into the hydrologic cycle, and to examine other human impacts on the climate and environment.

These activities sit alongside ANSTO's long-standing roles in medical research— contributing to new ways of treating conditions like Alzheimer's and Parkinson's diseases, as well as producing 85 per cent of the nuclear medicines used in Australian hospitals—and materials engineering, such as the development of Synroc—tailored ceramic forms for locking up high-level radioactive waste. ∎

SAVING SPECIES WITH BIG DATA AND APPLIED STATISTICS

Kerrie Mengersen, Professor of Statistics at the Queensland University of Technology (QUT), knows she has the best job in the world. Studying everything from the effect of chemotherapy on the brain to the efficient management of airports, Mengersen applies the tools of Bayesian statistics to some of the world's most important problems.

A Bayesian statistician views the world as an opportunity for adaptive learning, Mengersen explains. "We create new mathematical and statistical methods and computational solutions," she says. "Then we apply these new approaches to help solve important problems across diverse areas including health, environment, conservation, business and industry."

"As statisticians we find the stories in the data, and then convey these insights to people who make decisions to address some of the world's grand challenges."

Recently, Mengersen and co-authors caused a stir when they used this approach to study the number of living species that exist on coral reefs around the world. They found that despite decades of study, scientists appear no closer to understanding how many species are alive today. Existing estimates vary wildly, seemingly with little reference to previous findings.

They called for a more systematic and statistically based approach to this important topic. It's work that illustrates how statistical analysis can not only provide solutions, but also raise questions that can motivate discussion and generate new insights.

For Arun Sharma, QUT's Deputy Vice-Chancellor, Research and Commercialisation, Mengersen's work epitomises the university's focus on supporting high-impact research with a view to achieving significant public and practical benefit. "We're known as 'a university for the real world' because of our close links with end users," he says.

In 2013, the success of that strategy saw QUT ranked by The Times Higher Education as Australia's top university under 50 years in its 100 under 50 ranking. In recent years, QUT has built critical mass in the fields of robotics and automation, big data and personalisation—technological trends that are advancing disciplines, transforming professions and disrupting business models. These technological capabilities, coupled with an interdisciplinary culture and a desire to solve real-world problems, differentiate QUT's research.

The diversity and importance of Mengersen's work have led her to be named one of QUT's most successful researchers. In the past year alone, she was an integral leader in the consortium securing $20 million in funding over the next seven years to establish a new Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers of Big Data, Big Models, New Insights.

The aim of the Centre is to bring together researchers in mathematics, statistics and machine learning to create innovative models and make vital contributions to society, business and government. Its work is organised into three research programs—*Big Data Analytics* will focus on innovative methods for extracting and communicating information from big and complex data; *Big Models* will study new theoretical and methodological approaches to modelling large, complex systems; and *New Insights* will consider new ways of using the modelling and analytic approaches to inform our understanding of issues in the areas of health, the environment and societies.

"All our work involves engaging with great researchers and the expert clinicians, conservationists, managers and others who are doing amazing things in their own fields," Mengersen says. "It is humbling and very rewarding to be able to help them do their jobs better."

"One of the great things about working at QUT for me is its very strong support for maths and statistics," she adds. "The University encourages cross-disciplinary collaboration, which means that I actively work with people from social science, engineering, environmental science, health and business."

On top of all that, Mengersen is particularly proud of her group of researchers and students, known as the Bayesian Research and Applications Group, or BRAG. "What more could I ask of a workplace?" ∎

www.qut.edu.au
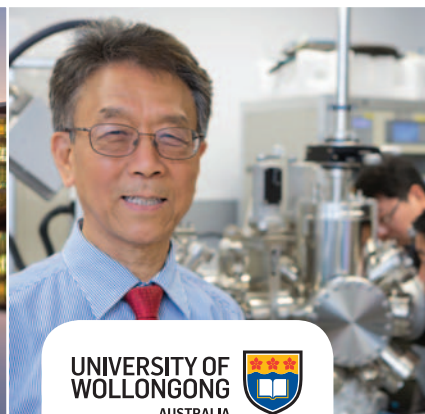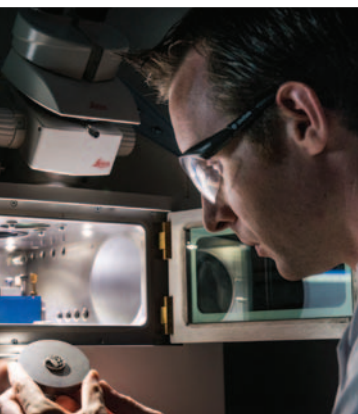
FROM HEAVY INDUSTRY TO NOVEL MANUFACTURING AND SUSTAINABLE SEAS

We live in an age of uncertainty, a world with a population that is both ageing and growing. It is a world of climate change, of rapid economic transformation and of immense technological advances.

Most cities and regions have developed slowly—over many decades, if not centuries. But in order to survive in this fast-changing world, communities must now be able to adapt quickly.

So says Professor Chris Gibson, leader of the Global Challenges Program at the University of Wollongong, one of a small number of universities around the world that is harnessing research strengths to target the global challenges of the 21st century.

TRANSFORMING LIVES AND REGIONS

Ranking in the top 2% of universities worldwide, we are especially qualified to tackle these challenges having played a leading role in the transformation of our home town of Wollongong, one hour south of Sydney, to a city of the future.

Throughout the last century, Wollongong was Australia's heavy industry heartland, best known as the home of the nation's largest steel mill. The University was established to train the industrial chemists, metallurgists and engineers needed to keep these factories working.

But, although these traditions continue, our research is now helping turn Wollongong into a 21st century centre for technology. This includes innovations like the BioPen, which lets surgeons "draw" on damaged bone with an "ink" containing live cells.

Professor Chris Gibson, Global Challenges Program Leader, University of Wollongong

Developed at the Australian Research Council Centre of Excellence for Electromaterials Science (ACES), which is headquartered at the University, the BioPen uses techniques similar to 3D printing to extrude cellular material between layers of gel. By delivering this material directly to the site of an injury, it can accelerate the regeneration of functional bone and cartilage.

The University of Wollongong was the first in Australia—and remains one of only five labs in the world—to fabricate silicene, a new two-dimensional material. Structurally similar to graphene, but made of silicon instead of carbon, silicene has intrinsic semiconductor properties that graphene lacks.

These properties could help it revolutionise materials science, leading to smaller, faster computer chips and more practical and efficient solar cells, as well as improvements in medical technologies and vehicle and aircraft parts.

Having demonstrated our expertise through achievements like these, we are now focusing our efforts on three Global Challenges—areas in which we can make the greatest impact: the ageing population, the growing pressure on marine and

coastal environments, and the changing nature of manufacturing and industry.

"We have taken an innovative approach to our research, concentrating on three challenges and then marshalling multidisciplinary resources and expertise from across the University to work on research projects under each challenge," Professor Gibson says.

MANUFACTURING INNOVATION

Taking advantage of Wollongong's history as a region of industrial creativity, we're already tackling the puzzle of what we should be making in Australia and how we should be making it. This means exploring the potential of the broadband-powered digital economy and the possibilities of smart materials, robotics and automation, as well as innovative medical devices.

It's also where we use our strengths in 3D printing and additive manufacturing to create results that were previously not feasible—such as custom-printed flutes that can play microtonal scales, or the notes between notes.

After bringing together economists, planners, social marketers, creative artists and designers to work on projects, we have initiated a global benchmarking exercise to ensure we are operating at best practice in these novel manufacturing techniques in our local region.

SUSTAINING COASTAL AND MARINE ZONES

Through our international collaborations we're seeking to improve the sustainable management of fish stocks, like bigeye and yellowfin tuna, that are under pressure from fishing fleets. This project, run through the Australian National Centre for

Ocean Resources and Security, will help to deliver the food security needed for social stability in Pacific island nations that face declining fisheries.

We're also harnessing social media technology to build resilience to climate change-induced extreme weather events in South-East Asian megacities. PetaJakarta uses real-time data from 150,000 Twitter users in Indonesia's capital to track and analyse flooding and inform emergency services and residents.

Our SMART Infrastructure Facility, which runs the project, was one of only six research institutions in the world to receive an inaugural Twitter Data Grant and gain access to this vital information.

LIVING WELL, LONGER

By taking a holistic approach to the problem of an ageing population—considering physical and mental health as well as access to services—we're discovering the physical and mental requirements for a long, healthy and high-quality life.

Our research is making a real impact on dementia, which is expected to affect one million Australians by the middle of the century. Two dementia-friendly communities that we're designing and piloting will not only accommodate but also actually welcome people who suffer from this debilitating illness.

And we're bringing together doctors, psychologists, geneticists, paediatricians, obstetricians and social scientists to follow three generations of local residents to understand their patterns of mental health and wellbeing as they age.

Professor Gibson says these collaborations capture the power of the University's multidisciplinary research capacity.

"Regions all over the world are facing common challenges, and it is important that from an empirical basis we develop tailored outcomes that have a global impact," he says.

The Global Challenges Program shows how we're playing a significant role in transforming lives and regions for the better.

To learn more, visit **globalchallenges. uow.edu.au**.

UNIVERSITY OF WOLLONGONG

A dynamic Australian university with a strong research focus, located just an hour from Sydney's International Airport.

CURRENT RANKINGS

- Top 2% of world universities (The Times Higher Education, QS, Leiden and Academic Ranking of World Universities)
- Top 100 for global graduates (QS Graduate Employers Survey)
- Five stars (QS World University Rankings)
- 1st in Australia for Educational Experience and Graduate Outcomes (Good Universities Guide)
- 22nd and 33rd in The QS and Times Higher Education rankings respectively for the world's Top 100 Universities under 50 years of age.
- 2nd (relative to size) and 8th on total research funding awarded nationally.

CAMPUSES

- Two campuses each in Wollongong and Sydney, and regional campuses in the Shoalhaven, Batemans Bay, Bega and Moss Vale
- University of Wollongong in Dubai.

KEY NUMBERS

- 31,500 students
- 12,800 international students
- 1,520 higher degree research students (893 domestic; 627 international)
- 138 nationalities represented in student body
- 2,350 staff
- 476 degree courses.

RESEARCH AND COLLABORATION OPPORTUNITIES

- Record $49.1 million in Australian Research Council (ARC) funding announced in 2013
- Postgraduate scholarships
- Multidisciplinary PhDs
- Vice-Chancellor's Postdoctoral Fellowships
- Visiting academic appointments
- Study leave and sabbaticals
- Financial assistance for international exchanges
- Joint PhDs with overseas universities.

LEARN MORE

Visit: www.uow.edu.au
         www.uow.edu.au/research
         Facebook.com/UOW
         Twitter.com/UOW

Phone: +61 2 4221 3555

UNIVERSITY OF WOLLONGONG
AUSTRALIA

University of Wollongong

RMIT UNIVERSITY: SMARTER TECHNOLOGY SOLUTIONS

With an international reputation for excellence in applied and outcome-oriented research, RMIT University in Melbourne, Australia, is a global university of technology and design and one of Australia's leading educational institutions.

Founded in 1887, RMIT has a proud history of strong links with industry. Under the leadership of RMIT Vice-Chancellor and President, Professor Margaret Gardner AO, the university has continued to foster connections between research and the private sector. At present, RMIT has over 200 active research collaborations with industry and overseas partners.

With a student body that includes more than 10,000 postgraduate students, 5,000 staff comprising 2,500 academic and teaching staff, and campuses in Australia and Vietnam — plus a recently opened center in Barcelona, Spain — RMIT boasts a rich portfolio of teaching and research partnerships that cover every continent.

As part of its strategic plan for 2015, RMIT has set itself three goals: to be global, urban and connected. Entwined with this is the university's desire to develop innovative solutions to tackle the complex, technical challenges of today. In particular, researchers at RMIT are developing smart-technology solutions in the areas of energy, health and computing.

SMART SOLAR SOLUTIONS

Solar, or photovoltaic, cells are a common feature in today's cities and the electricity they generate from sunlight can make a significant contribution to meeting everyday power requirements. An even simpler idea is to take sunlight and convert it directly into heat. Gary Rosengarten (pictured, top left), professor of sustainable systems engineering at RMIT's School of Aerospace, Mechanical and Manufacturing Engineering, is doing just that — by applying smart technology solutions to the field of solar thermal energy.

"Solar thermal energy is fundamentally more efficient than photovoltaics," explains Rosengarten. While photovoltaics struggle to reach 20 per cent sunlight conversion efficiency, solar thermal energy can easily achieve 70 per cent efficiency. Plus, heat is also easier and cheaper to store than electricity.

Rosengarten leads the MUSIC (Micro Urban Solar Integrated Concentrators) project, which aims to revolutionize the use of solar collectors in urban environments. The thin and lightweight concentrators being developed will share the look and placement of conventional photovoltaic panels. However, by using a clever combination of vacuum insulation to keep the heat in and mirrors or lenses to concentrate the light, the concentrators will be able to reach temperatures of up to 400 degrees Celsius, without the need for expensive systems that track the Sun.

Another of Rosengarten's projects combines solar thermal and photovoltaic technologies to efficiently create both electricity and hot water. This system employs a collector to appropriately siphon off solar wavelengths that can be efficiently absorbed by photovoltaic cells, allowing the remaining light to reach a solar thermal collector, generating temperatures of up to 150 degrees Celsius.

For Rosengarten, RMIT's mission of tackling key issues that affect communities and the environment is important. "I like to know that I am contributing to society by helping industry develop new products for market and that I am doing it in an environmentally sustainable way." He is also impressed by the atmosphere that the university has created for its researchers. "RMIT has invested considerably in people and infrastructure in targeted areas to ensure there is critical mass to achieve world-class outcomes," he says.

CONNECTING MEDICINE, PHYSICS AND ENGINEERING

RMIT's location in the city of Melbourne places the university within easy reach of 19 major research hospitals and facilities, meaning researchers are ideally positioned for immersion in the medical community. Consequently, smart technology solutions to meet biomedical needs are constantly evolving at RMIT, as are the facilities in which they are being developed.

"At RMIT, world-class facilities are being set up and a growing group of leading researchers are working together to publish their research in leading journals," says James Friend (pictured, top middle), director of the flagship AUS$30 million MicroNano Research Facility (MNRF) at RMIT and a professor at the RMIT School of Electrical and Computer Engineering.

2014 will see the completion of the MNRF, a significant upgrade of RMIT's existing fabrication and metrology capabilities. The new facility will meet the needs of research at the nanoscale and integrate biomedical research activities. Research at the facility will span the traditional disciplines of physics, chemistry, engineering and medicine.

Much of Friend's own research lies at the boundaries of medicine, physics and engineering, and his development of miniaturized motors, or microactuators, for the retrieval of blood clots from deep within the brain is a prominent example of this union. The use of tiny actuators capable of navigating weakened blood vessels will enable minimally invasive neurological intervention in people affected by strokes or aneurysms, who might otherwise be unsuited to treatment.

Friend's background in microfluidics and micro- and nanofabrication is also contributing to substantial improvements in the delivery of drugs via the lungs. By employing acoustic waves that travel along a material's surface, his group can atomize large biomolecules — including drugs, DNA and antibodies — and even cells into suitably sized droplets, while avoiding the damage that conventional methods of nebulization can cause.

The technology has already proven effective for the delivery of a DNA vaccine in sheep, and has the potential to be used to convey gene therapy or stem cells to inaccessible sites, deep within the lungs. "Our work promises to overcome fundamental problems neurosurgery, drug delivery and microfluidics," explains Friend. "This is a huge motivation for conducting my research."

COMPUTING WITH DIAMONDS

RMIT is a university focused on both design and technology. For Andrew Greentree (pictured, top left), an associate professor at the RMIT School of Applied Sciences, "it is the technology part that is key".

Greentree is guiding RMIT's efforts to bring computing into the quantum age. A quantum computer stores information as qubits — the equivalent of bits in classical computing. But unlike classical bits, qubits can exist simultaneously in multiple states, a feature that can be exploited to allow new and more powerful forms of computing. And while a classical computer uses only one aspect of quantum mechanics — barrier tunneling — a quantum computer also uses the quantum phenomena of superposition and entanglement to harness the power of quantum states. "This provides greater control, allowing computation to be carried out in new ways," observes Greentree. "Simply put, a quantum computer is the ultimate computer."

For information to be transmitted, qubits must travel from one location to another. Although wires serve this purpose in classical computing, they are incapable of preserving quantum information. Before joining RMIT, Greentree conceived an elegant solution to this problem, called coherent tunneling adiabatic passage (CTAP), which continues to inform his current research. CTAP will likely play a key role in facilitating the on-chip transport of quantum information.

At present, quantum computing devices that employ a small number of qubits already exist. But scaling the technology up remains a challenge — albeit one that Greentree is willing to embrace. "I want to ensure that my research is important and relevant," he says, "and to transition my research into real world applications and devices."

Greentree is also an expert in the use of diamond for quantum purposes. In particular, he hopes to use diamond to create hybrid quantum–classical computers that boast solid-state quantum memories. Beyond computing, Greentree's interests in diamond are shaping part of the research programme at the upcoming Australian Research Council Centre for Nanoscale BioPhotonics, in which RMIT is a partner institution.

RMIT is a university that is excited by change and isn't afraid to back key ideas from its researchers, he notes. "If quantum technology is as important as we think, soon we will need to be educating quantum engineers who are ready to design and build practical quantum devices," he says. "It's my aim that RMIT will be at the forefront of this revolution."

The University of Auckland
MAKING OUR MARK
IN LASER SCIENCE AND OPTICS

Information processing and advanced telecommunications technology are just some of the disciplines that laser science and optics have the potential to impact in the future.

Temporal cavity solitons – as persistent light pulses – demonstrate the beauty of laser science and constitute the most fundamental example of self-organisation in optics. They are an ideal information carrier to store data and create an all-optical buffer, a critical function for high-speed routing technology.

University of Auckland Associate Professor Stéphane Coen has contributed to the first-ever capture of these pulses of light using nothing more than a continuous-wave laser and an unamplified loop of standard optical fibre. Now his team at Auckland can retain pulses for more than one hour, equivalent to a propagation distance of close to one billion kilometres, and has evidenced their ultra-weak long range interactions. These pioneering achievements have implications for many disciplines, from life sciences to ocean physics.

Published in Nature Photonics, the research has revealed the rich physics of these little-known objects. It also has deep practical significance, as temporal cavity solitons provide key insights into new microscopic light sources that have the potential to provide ultra-accurate measurements on a chip.

Associate Professor Stéphane Coen is one of a team of award-winning scientists achieving international recognition for their work in our world-class laser laboratory.
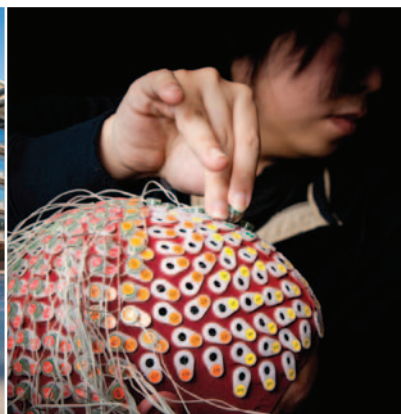
To find out more, visit
LEADINGRESEARCH.ac.nz

THE UNIVERSITY
OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

KingSt 12580/NAT/A

University of Western Sydney

UNIVERSITY OF WESTERN SYDNEY: TACKLING BIGGER, UNANSWERED QUESTIONS

Experimental psychologists, neuroscientists, engineers, computer scientists, linguists, composers and dancers — few people might guess what kind of research projects would require such a diverse team of professionals, but Kate Stevens from the MARCS Institute of the University of Western Sydney (UWS) knows. She is the leader of Music Cognition and Action, a research program that uses behavioral and neurophysiological approaches to study the temporal dynamics of perception and creativity, as well as individual and group performance.

"We are trying to understand how innovation and knowledge emerge from a group," says Stevens, "However, these are difficult and complex questions that require an interdisciplinary team to answer."

Stevens has joined forces with a professional dance company and built a research team of internationally recognized scientists, choreographers and dancers. They analyze behavioural and brain processes as people interact and create. Every week, Stevens holds seminars and specialist group meetings where other research groups at the institute join and contribute to the discussion. "Here at the MARCS Institute we have experts in, for example, perception, language development, neuromorphic engineering and human-machine interaction," says Stevens. "We all come together for one purpose: to answer those big and risky questions."

FOCUSSING ON IMPACT

Scott Holmes, the Deputy Vice-Chancellor of Research and Development at UWS, says Stevens' story is reflective of the university's emphasis on research excellence. "Our big goal is to make sure all of our research is high impact, not only in terms of citations but also the effects on professions and our communities," he says.

Indeed, UWS has come a long way since its founding in 1989. Over the past 25 years, the institution has evolved into a vibrant, modern, outer-metropolitan university where students receive professional training in science, health and medicine, technology, engineering, humanities, education, business and law. UWS is now home to over 3,000 staff and 42,000 students, most of whom are residents of Western Sydney — one of the fastest growing and most culturally diverse regions in Australia.

UWS places a strong emphasis on research and development. It has invested heavily in areas including neuroscience, infrastructure engineering, complementary medicine, education and humanities. 70% of UWS research reviewed by the Excellence in Research for Australia assessment in 2012 was rated as "world standard" or above.

BIG INFRASTRUCTURE, BIG ANSWERS

Environmental sustainability is a core area for UWS. The university has invested heavily in the Hawkesbury Institute for the Environment, which focuses on the impacts of environmental change on terrestrial ecosystems. "Our research spans from the molecular and microbe level through to plants, animals and entire ecosystems," says Ian Anderson, Director of the institute. "The findings from our research are aimed towards helping us better manage both Australia's native forests and intensively managed landscapes in the future."

One of the largest projects conducted at the Hawkesbury Institute for the Environment is EucFACE, a free-air CO_2 enrichment field experiment that examines the response of a native Australian eucalypt forest to an atmospheric CO_2 concentration of 550 ppm, which is expected later this century. "We really stand out internationally in terms of our research programs and our extensive array of world-class research infrastructure to study climate change impacts," says Anderson. "We run a lot of large-scale, comprehensive experiments and the university has been very supportive."

While EucFACE is still in its infancy, preliminary findings indicate that mature eucalyptus trees may not respond to future concentrations of CO_2 as predicted by findings from experiments in other parts of the world. "A significant difference is the poor nutrient status of most Australian soils and the fact we have very long dry spells where there is not a lot of water. So the ability of these tree species to respond to increasing CO_2 level is limited," says Anderson.

University of Western Sydney

University of Western Sydney
www.uws.edu.au
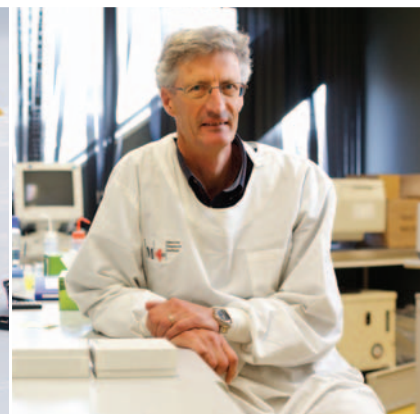
AN ISLAND LABORATORY HELPING TO SAVE LIVES AND SPECIES ACROSS THE WORLD

The University of Tasmania has a growing reputation as one of Australia's foremost teaching and research institutions, ranking in the top two per cent of universities worldwide and in the Australian top ten, and leading the world in our distinctive research themes.

While maintaining our island identity, we undertake research of international scope, thanks to collaborative partnerships we've built across the globe, strengthened by a network of more than 90,000 alumni spanning more than 120 countries.

BETTER HEALTH: FOR BABIES AND FOR GROWN-UPS

Beginning in the late 1980s, University researchers investigated why Tasmanian babies had a relatively high risk of sudden infant death syndrome (SIDS). Their discovery that sleeping on the stomach is a major risk factor saves the lives of hundreds of babies every year.

This was one of the first of many successes of what is now the Menzies Research Institute Tasmania. Recently, its scientists have discovered genetic markers for prostate cancer risk, and how vitamin D helps prevent multiple sclerosis relapse.

ENVIRONMENT, RESOURCES AND SUSTAINABILITY: DEFEATING DEVIL'S DISEASE

Medical scientists from the Menzies Research Institute Tasmania are helping our zoologists study the facial tumour disease that's devastating Tasmanian devil populations. They've found a genetic mutation in the contagious cancer that makes it invisible to the devil's immune system, meaning that cancer cells pass to a new devil without triggering a protective response.

This discovery raises hope for a future vaccine, but in the meantime we are protecting populations of healthy devils on isolated islands and peninsulas, as well as in captive breeding facilities throughout Australia.

CREATIVITY, CULTURE AND SOCIETY: UNDERSTANDING CONVICT LIFE

University historians gained new perspectives on Australia's convict history and other aspects of colonial life from records of the 73,000 people transported to Van Diemen's Land, gathered with the help of Oxford, Sussex and Liverpool universities.

"It's about using Tasmania as an island laboratory to look at national and international problems," says Hamish Maxwell-Stewart, leader of the Founders and Survivors project.

These insights have led curators of heritage sites like Port Arthur to replace stereotypical blood-soaked interpretations with a more complete picture of colonial life.

MARINE, ANTARCTIC AND MARITIME: PRESERVING FUTURE FISH

Marine parks must be more than just boundaries on a map if they're to conserve biodiversity, our researchers found with the help of recreational divers. Their survey of 87 marine protected areas in 40 countries found that many were no more diverse than non-protected areas nearby.

Most successful were parks with a well-enforced ban on fishing; also those more than 10 years old, relatively large and some distance from fished areas.

"It's these kinds of areas that we need to create and at the same time retrofit the existing ones that are unlikely to ever reach their conservation goals," said the study's lead author, Graham Edgar.

ENABLING TECHNOLOGIES AND PLATFORMS: BACKWARDS IN TIME

For a billion years of Earth's history, life stopped evolving and remained little more than a layer of slime. University geologists used a technique developed for studying mineral ores to determine that the pause was due to a shortage of oxygen and other elements essential for life.

The team at the Centre for Ore Deposit and Exploration Science vaporised ancient mineral samples and measured the levels of trace elements present when they formed.

"The fluctuations in these levels may help explain events like the emergence of life, mass extinctions, and the development of gold and other ore deposits," says geologist Ross Large.

Our publication Research to Reality has more stories of research from University of Tasmania, at: www.utas.edu.au/research-to-reality. ◼

AUSTRALIA
ECU
EDITH COWAN UNIVERSITY
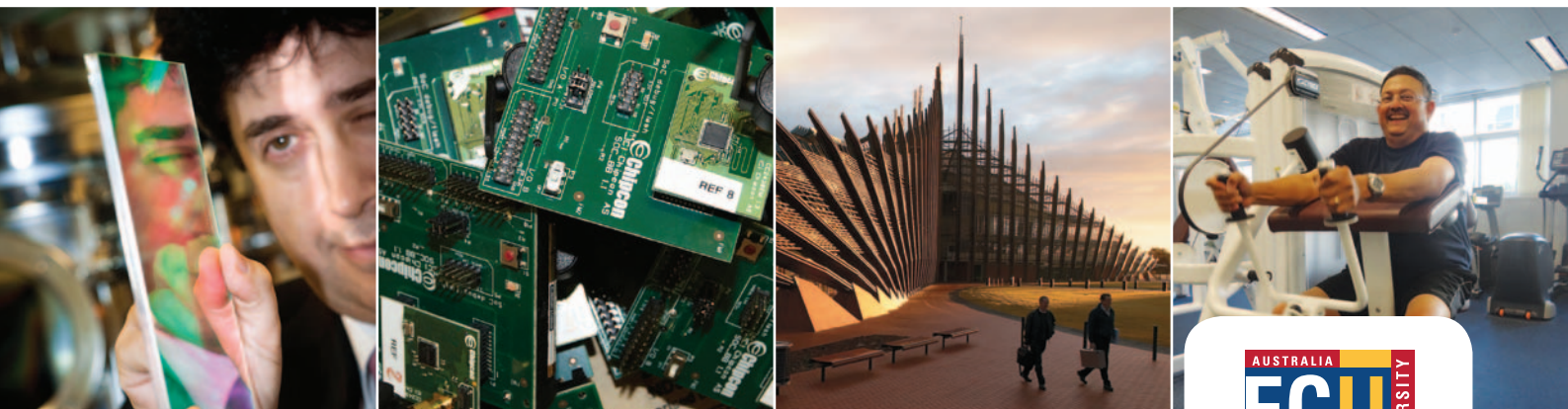
SMART IDEAS FOR CANCER RECOVERY, SAVING ENERGY AND FIGHTING CRIME

Edith Cowan University in Perth, Western Australia, is recognised for research that tackles real problems, concentrating our efforts in areas where we have a proven track record.

Our research institutes—including Health and Wellness, Electron Science Research and Security Research—and our close links with nine other Australian universities are part of our commitment to a vibrant research culture and strong support for our staff and student researchers.

WORKING OUT BETTER LIFE FOR CANCER PATIENTS

Toxicity in cancer treatment leads to a number of well-established musculo-skeletal deficits, as well as osteoporosis and bone fractures that can substantially reduce quality of life, physical function and independence. But research by the University's Health and Wellness Institute has shown that exercise can act as medicine to significantly reduce these risks and help cancer survivors recover.

Their studies on prostate cancer patients found clinical benefits of resistance training for reversing muscle loss and improving physical function and quality of life.

This approach is a big change from historical practice, in which clinicians advised cancer patients to rest and avoid activity. However, exercise guidelines co-authored by the Institute have now been adopted as the prescribed model for cancer management in North America and much of the world.

WINDOWS FOR A CLEAN ENERGY FUTURE

The vast area taken up by windows on buildings could soon become a source of electricity, by using technology that turns them into transparent solar panels. TropiGlas, being developed commercially by a Perth company using research from the University's Electron Science Research Institute, is an "intelligent glass" technology that converts the ultraviolet and infrared components of sunlight into electricity while allowing visible light to pass through.

A hair-breadth, nano-engineered film sandwiched between two sheets of glass selectively diverts the energy of this radiation to photovoltaic cells at the edge of the pane. Although not yet as efficient as traditional solar panels, TropiGlas is able to cover a much larger area than a rooftop installation, effectively turning buildings into power plants.

"We believe that within a couple of years we should be able to double the efficiency, which is very good considering we're not using visible light," says Institute director Kamal Alameh.

The first commercial trials of the product are taking place in South Africa, where TropiGlas is being installed in a new government office building in Pretoria.

SIMPLE SOFTWARE FOR CYBER SLEUTHING

Some of the most serious modern crimes are also the most technically challenging to police, requiring the analysis of suspects' computers for illegal images such as child pornography. But the University's Security Research Institute has produced software that lets police view photos and videos on hard drives without damaging their integrity as forensic evidence.

Called SImPLE—short for Simple Image Preview in Live Environment—the software was designed to be both effective and easy for officers to use with only minimal training. This contrasts with the currently used sophisticated, multi-purpose software, which can be operated only by a central unit of expert technicians.

Doing the analysis on-site also enables a quicker response for identifying and rescuing victims of abuse, as well as providing evidence needed to bring a suspect into custody or elicit a confession.

In addition to harnessing the expertise of the Institute's researchers, the development of SImPLE used the skills of University students, giving them the chance to contribute to a project with substantial community benefits. As the project moves into its beta stage, the students will continue to support the software and keep it up to date in a rapidly changing field.

Find out more about our work in all fields at **www.ecu.edu.au/research**. ▪
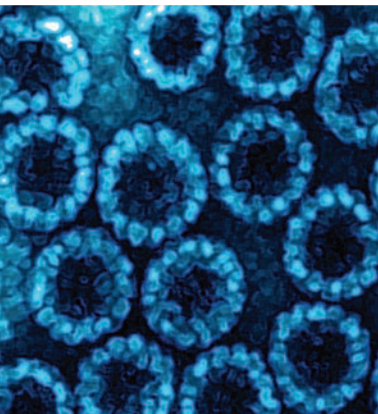
KNOWLEDGE LEADERSHIP FOR A BETTER WORLD

Ranked in the world's top 100, The University of Queensland (UQ) is one of Australia's leading teaching and research universities, and has educated more than 210,000 alumni—including over 11,000 PhDs—who have made and continue to make positive impacts throughout the world.

UQ's excellence in research has been translated into positive impact for people and communities worldwide, with many economic, environmental, health and social benefits. Leading outcomes include:

♦ the world's first vaccine against cervical cancer, which has benefited tens of millions of women worldwide, and has the capacity to save an estimated quarter of a million lives annually

♦ the Triple P – Positive Parenting Program, which has reached more than seven million children and their families in approximately 25 countries, and has been translated into more than 20 languages

♦ signal correction technology used in approximately two-thirds of the world's magnetic resonance imaging (MRI) machines

♦ genome mapping studies that have greatly increased our understanding of the genetic causes of chronic, debilitating conditions such as ankylosing spondylitis, osteoporosis, tuberculosis, leukaemia and rheumatoid arthritis

♦ mine safety technology (GroundProbe) used to preserve lives in some of the world's largest mining companies as best practice for active slope monitoring

♦ conservation planning software (Marxan) used in more than 100 countries to support the design of marine and terrestrial reserves

♦ innovative treatments for diseases linked to autoimmune, neuropathic and inflammatory conditions, which are being translated by UQ biotechnology start-ups including Spinifex and Dendright.

To bring these and many other innovations to the global market, UQ's excellence in research has been enhanced by the University's commercialisation companies: UniQuest, which ranks in the world's top 10 per cent of such companies, and JKTech, which specialises in commercialising resource industry-related innovations.

This translation of research excellence to real impact for society defines the "excellence-plus" culture embedded at UQ. We particularly seek to partner with industry, government and non-government organisations to enhance these outcomes. In 2012, UQ received more research funding from non-government sources than any other Australian university, and UQ consistently features among the top few Australian universities for research funding from government.

UQ's success is underpinned by the quality of its people and research excellence, which is evidenced by the Australian Government's Excellence in Research for Australia (ERA) 2012

Research conducted at UQ by Professor Ian Frazer AC (pictured) and the late Dr Jian Zhou on virus-like particles led to the development of the world's first cervical cancer vaccine.

assessment. In this benchmarking exercise, all of UQ's research fields were ranked as world standard or above.

With a focus on excellence in discovery, learning and engagement, UQ's record of success has far-reaching impacts. We are committed to "excellence-plus" to enhance UQ's positive worldwide impact. ■

THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

www.uq.edu.au

TOP RIGHT IMAGE COURTESY OF GROUNDPROBE ®

Advertiser retains sole responsibility for content

Achieving positive impact together

At CSIRO we shape the future. We do this by using science to solve real issues. Our research makes a difference to industry, people and the planet.

Connect with us:

An open letter from Dr Megan Clark, Chief Executive CSIRO

Dear valued members of the international research community,

Responding to the major scientific challenges of our time takes an agile and considered approach, as well as the ability to change and evolve.

Indeed, embracing change is essential for CSIRO to continue to be one of the top ten applied research agencies in the world and Australia's most trusted name in science and technology.

In order to make it easier to achieve common goals for long-lasting benefit we have streamlined CSIRO into nine National Research Flagships.

These National Research Flagships allow us to focus on the biggest challenges that face our nation and the globe. By taking a multidisciplinary approach, we can bring to bear the expertise we need to undertake groundbreaking scientific research.

We're creating an operating environment where we can work with you, the international research community, to create positive impact within and across our areas of specialty, and on the world around us.

This is an exciting time for our research. Our doors remain open for researcher exchanges and visits, joint ventures, collaboration agreements, joint publications, joint forums for knowledge exchange and co-investment. We want the best minds in Australia working with the best minds in the world to solve the greatest challenges of our times.

We look forward to continuing to make a positive impact together.

Dr Megan Clark

ACHIEVING SUCCESS IN COOPERATIVE RESEARCH CENTRES

As Australia's University of Enterprise, the University of South Australia (UniSA) engages meaningfully and creatively with our peers, industry and the community in areas that are aligned with key state, national and international priorities.

We are proud of our collaborative and solutions-focused research through our successful involvement in the Commonwealth's Cooperative Research Centres (CRC) scheme, where we currently lead and headquarter three CRC's and partner in eight more.

CRC's in the areas of Contamination Assessment and Remediation of the Environment (CRC CARE), Cell Therapy Manufacturing (CTM CRC) and Data to Decision (D2D CRC) place UniSA at the forefront of industry and academic research collaboration, as we bring together our breadth of research expertise to focus on end-to-end approaches and solutions.

From areas of cancer, through to energy and the environment and the development of future industries, our enterprising research is delivering rich, robust and multifaceted outcomes that make a difference to the world.

To learn more about our collaborative research success, please visit **unisa.edu.au**

University of South Australia Senior Research Fellow Dr Mohammad Mahmudur Rahman is performing CRC CARE-supported research on arsenic toxicity, bioavailability and detoxification. This work will help minimise the risk of arsenic exposure at contaminated sites such as disused mines.

University of South Australia